

London Zoo crippled by misjudgement

London will lose its city-centre zoo in September if last week's flaccid announcement from the London Zoological Society is to be believed. This is a sad ending to a mostly honourable tale.

THE Zoological Society of London's announcement last week that its zoo in central London is to close next September is innocent of all signs of awareness of the enormity of the proposal, and of the irony that it should be closing one of the world's most accessible zoos at a time when public consciousness of the diversity of species has hardly ever been greater. It is as if the operators of a particle accelerator had decided to go out of business just when they were about to find the *top* quark, or some other exotic particle (see below). But there is more than black irony in what the London Zoo proposes. Through much of its history, the zoo has been managed successfully by a scientific society, whose reputation has been enhanced on that account. By the same test, the society's failure to make sense of its public responsibilities over the past five or ten years will reflect badly on its standing, as well as on the reputation for competence of other scientific societies in Britain.

In the circumstances, the comment by Lord Zuckerman on page 621 is remarkably restrained. For one who has spent a substantial part of his life's energy on the welfare of the London Zoo to see it collapse in the hands of others must be more galling than he allows. Could that collapse have been averted? The muck-raking question cannot be avoided. The Zoological Society's crucial mistake was the particular way in which it chose to use an unprecedented government grant of £10 million (in 1988) to make a dash for commercial viability. Instead of building on its own homespun strengths, and building a constituency of public support for the London Zoo, the society delegated the execution of its plans to people with an entrepreneurial cast of mind. When, just over a year ago, the money started running out, their response was to go back to the government for more against the drumbeat of a public relations campaign in which named members of endangered species were declared to be in danger individually. It was a rotten way to run a serious zoo.

But may that not have been inevitable in the closing years of Mrs Margaret (now Lady) Thatcher's era? The fashion of the times was that the crucial test of the virtue of an enterprise is its capacity to survive in the marketplace. Many British research laboratories were closed because they could not satisfy the test. (Some of those that could were promptly sold to private owners.) The trouble at the London Zoo is that it has priced itself out of the market; its admission prices are so high that its mass audience (especially among children) has been frozen out, but even the resulting revenue stream has not kept pace with mounting costs. If the London Zoo,

instead of bending its energies to the sale of expensive admission tickets, had done more to sell zoology as such, it might well have acquired the stream of subscription revenue (from local authorities and private persons) on which other zoos depend. The dash for viability was a marketing exercise in favour of the wrong products.

Just where the dust will settle in the next few months is anybody's guess. The Zoological Society plans that Whipsnade Zoo should continue. It also hopes that its in-house research institutes will continue, unaffected by the collapse of its main zoo. But that betokens a certain optimism. The institutes are at present supported by grants from the Universities Funding Council, channelled through the University of London. Their research is an interesting blend of conservation topics, animal physiology and molecular taxonomy. But the closing of the London Zoo is certain to draw attention to the anomalous status of the institutes, which would be better placed under the umbrella of the research councils. It is to be hoped that that transition will be carried out in a seemly fashion.

Whether animals will finally disappear from the site of the London Zoo in Regent's Park is also, at this stage, uncertain. A scheme canvassed by private developers for replacing the zoo by a conservation exhibit should not be scorned. Whether, at this late stage, it would be possible to rescue the zoo in its present or some similar form seems much less likely: there is precious little time between now and September. But even that apparently wild hope should not be dismissed. To let decades of intellectual and financial investment in Regent's Park be thrown away because of the misjudgements of the Zoological Society will be a great misfortune. □

Another phoenix?

The SSC is not dead (yet), but the US Congress may yet force a truly international effort in high-energy physics.

HAS the United States turned its back on big or even adventurous science? That is the question widely prompted by the vote in the US House of Representatives last week to deny further support for the Superconducting Super Collider (SSC), except for the pittance required to close up shop. But the vote itself does not mean that the SSC is dead. Procedurally, the Senate must first decide on the administration's request for \$650 million for the financial year beginning on 1



October; then, if the House and Senate differ, there will have to be a conference from which a compromise may emerge. The SSC could yet limp into 1993 on a reduced budget. But that would not be widely welcomed.

The background to the decision implied by last week's vote is the general concern about the state of the US economy, and the particular need to contain the federal deficit. In a hard year, \$650 million of discretionary spending is a worthwhile prize for hard-pressed politicians. That is why too much of general significance should not be read into the decision. Thus Representative George Brown, the chairman of the House Science Committee, was off the mark when he described the vote as a decision that the United States "cannot afford the next level of understanding" in an important field of science. Those directly involved have mercifully been more phlegmatic. George Trilling, who leads a team building one of the two detectors for SSC, says that, in the absence of technical snags, the decision shows "how the political winds are blowing".

Even so, the decision is disconcerting, with implications outside the United States. What will happen now at the European High-Energy Physics Laboratory (CERN) in Geneva? By the end of this year, it had been hoped that the government members of the CERN consortium would have decided to embark on the construction of the Large Hadron Collider (LHC), the nearest competitor to the SSC in sight. Will the US decision stiffen the resolve of the CERN council? Or will it, alternatively, persuade CERN that the urgency of that project is not now so great?

That dilemma for CERN is also an opportunity for high-energy physics. For decades, in Europe, the erstwhile Soviet Union and the United States, high-energy physicists have acknowledged that genuine international collaboration will ultimately be required in the construction of particle accelerators, but have insisted that the time will come "not now, but later" — usually after their current machine is built. The best outcome of last week's vote in Washington would be an application by the United States to belong to CERN, which could then be renamed to acknowledge its extended membership. That would be the best way of making sure that the high-energy community, having probably lost the SSC, will not lose the LHC as well.

Sadly, though, there is a procedural snag, dramatized by last week's vote. For the past year, the United States has been pleading with (some would say bludgeoning) Japan to contribute to SSC. The Japanese government had promised an answer by the end of the year. After last week's demonstration of congressional fickleness, it is not now difficult to guess what Japan's answer will be. But that same fickleness is an impediment to US membership of CERN, which is an organization established by an international treaty that requires members to pay their dues on time. The procedural snag is that treaties duly ratified by the US Senate do not bind the Congress as a whole to vote prescribed funds. (US arrears in paying United Nations dues are a spectacular case.) Dr D. Allan Bromley, the US Presidential Science Advisor, has been promising a remedy since he took office two years ago. He had better hurry, or the uncertainty created by last week's

vote will quickly spread to overseas partners in other US ventures, the space station in particular. □

Leeds disunited

There is no substitute for openness in the handling of allegations of scientific misconduct.

THE University of Leeds has a lot to learn about allegations of fraud and their investigation. For more than a year, the university has been wrestling with a complaint by Dr Chris Chapman, an immunologist employed by the National Health Service at the university's teaching hospital, that academic colleagues with whom he had worked had been guilty of fraud. Chapman had been asked to develop an immunoassay for the lymphokine interleukin-6, the gene for which had supposedly been cloned by academic colleagues (crucially, employees of the university) and used *in vitro* to produce the protein. At some stage, Chapman says, one of his academic colleagues claimed that the biological activity of the preparation had been demonstrated *in vivo*, but that claim turned out to be false. The consequence was that several people worked for months on a pointless project.

That is not fraud in the usual sense of a deliberately misleading publication, but does qualify as scientific misconduct to be taken seriously. How did the university respond? After badgering from Chapman, the university set up an inquiry described, in a laconic statement last month, as "confidential". The statement also said that "remedial action" had been taken, but went on to refuse further discussion of the issues arising. Meanwhile, and quite separately, the hospital management at Leeds has been reappraising its need of qualified staff and has concluded that it has no further need of Chapman, who has been made redundant more than a decade before his normal retirement age.

This does not look well, to say the least of it. By now, there is an almost biblical stock of case-law to show that institutions can hush up allegations of misconduct only at their peril. In the long run, even those accused do not profit from the haze of rumour that inevitably takes the place of measured analysis and appraisal, as when inquiry reports are made public. On this occasion, the university itself will suffer by giving an open licence to those who would magnify in their imaginations the scandal on which it will not comment. The fact that the whistle-blower in the case has lost his job, however good the reasons advanced by the hospital management, makes matters look even worse.

In backwaters such as Britain, it is tempting to suppose that scientific misconduct, which is known to be well correlated with achievement, is comparatively rare on that account alone. But the correlation is not perfect; indeed, the temptation to cut corners may be just as great as in competitive environments, even if the ambitions of the corner-cutters are appropriately constrained. Not just Leeds but other British universities should be on the look-out for these minor transgressions in case their collective ambition to become good again should ever be attained. □

US company's gene therapy trial is first to bypass RAC for approval

Washington. Researchers at a US biotechnology company are expected to be the first group to conduct a gene-therapy experiment without receiving approval from the Recombinant DNA Advisory Committee (RAC), the government review committee set up more than 15 years ago to assuage public fears about experiments involving genetic manipulation. Last week, a committee of the US Food and Drug Administration (FDA) recommended that the agency allow Viagene Inc. of San Diego, California, to proceed with clinical trials of a gene therapy involving a human immunodeficiency virus (HIV) vaccine, the first time that such an experiment has been treated as a routine drug trial rather than a foray into unknown territory that requires extraordinary safety and ethics review.

Viagene was able to sidestep the usual RAC approval process because it is a private company and receives no federal funds. But with at least half a dozen similar companies preparing their own gene-therapy trials, this route is expected to become well-trodden as industry researchers seek to avoid the public hearings and scientific grilling of the RAC.

The FDA and the RAC differ in more than just name and methodology. In general, the RAC embodies much of the public concern of the late 1970s and early 1980s, and its 24-member panel includes lawyers, ethicists and political scientists as well as genetics researchers and clinicians. Its meetings are open, announced in advance and often attended by the media and those concerned about the use of genetic engineering techniques. The RAC has also taken a broad interpretation of its mandate to ensure that gene therapy protocols are safe and ethical.

To be ethical, in the RAC's view, an experiment must be conducted with all the appropriate patient consents and precautions and also have a reasonable chance of success. This generally requires researchers to demonstrate, with data from animal models and the like, that there is solid scientific evidence to suggest that the experiment will work in humans. For many groups, that has proved difficult. At a meeting earlier this month, the RAC rejected two protocols on the grounds that they had insufficient supporting data, and another protocol was withdrawn when it became clear that it, too, would be rejected for the same reasons.

The FDA, in contrast, is principally concerned about safety. For a Phase I trial — the very earliest of the three phases of approved clinical trials — the FDA generally wants to know only that the drug or therapy is not toxic and is not likely to harm patients. It relies on existing rules for obtaining proper

consent, and, unlike the RAC, it treats gene therapy no differently from other drugs and therapies in that regard. Unless a company asks the FDA to open its approval process, it is generally closed from public view.

But as different as the two approval routes are, last week's launch of the FDA approval



process resembled nothing more than the RAC itself. The review committee opened most of its deliberations to the public, and the FDA brought in six specialists to evaluate the scientific merits of the protocol. No one, however, expects the FDA to go to such extreme lengths for every protocol. In fact, most gene-therapy applications will probably be treated like any application for an experimental drug or therapy.

Viagene asked the FDA to conduct its review in public to remove suspicion that the company was trying to bypass the usual safeguards. "We felt that because it was the first one, we wanted it to be open", explains Steven Meto, vice president for research and development. But he predicts that "down the line, it may be more closed".

Meto says that he does not expect all industrial gene-therapy trials to go through the FDA. Some may have to be submitted to the RAC either because they are carried out with federal grants or because the institutions that do the clinical trials receive federal funds.

Another disincentive for companies is the length of the FDA process. Viagene, for example, spent a year and half developing some 3,000 pages of documentation just to get to Phase I clinical trial approval. Of course, all gene-therapy protocols must receive some sort of FDA approval, whether they go through the RAC or not, and if a

company intends to market a product it will eventually have to receive full FDA marketing approval. But the RAC, despite its public nature and its hunger for efficacy data, may be easier for those conducting basic research and for protocols that do not have a near-term therapeutic goal.

Some of those who called for the creation of the RAC are hoping that companies will continue to use the committee voluntarily. "We're concerned that as gene therapy gets more sophisticated, the process is getting less accountable," says Andrew Kimbrell, an attorney at the Foundation for Economic Trends, the umbrella organization for biotechnology gadfly Jeremy Rifkin. He notes that the FDA cannot match the in-house scientific expertise of the RAC. "I'm sure that the FDA is not the proper body to regulate what is involved", he says. "I'm very concerned about private companies using this route." FDA officials say that they will continue to use outside experts as necessary.

But in the face of increasing acceptance of gene therapy, those who still wish for special reviews of everything from the basic biology to the theology of simple gene-transfer experiments appear to be losing ground to those who argue for business as usual.

Christopher Anderson

OED on CD

London. What do the words 'heterogenist', 'condensational', 'subfraction' and 'zoarchaeologist' have in common? According to the *Oxford English Dictionary* (2nd edition), they all had their first recorded use — in 1870, 1903, 1946 and 1987 respectively — in *Nature*. This week's launching of the compact disc version of the dictionary will give philologists a chance to explore in ways not previously possible the half a million definitions and 2.4 million quotations contained in the printed edition.

Searches can now be made for headwords, phrases and variant forms, and the text can be perused according to date, quotation, pronunciation, definition, etymology and much more. For example, *Nature* has been cited 9,737 times for early use of such words as 'filter', 'urogastrone', 'abactinal' and 'stagnating'. In comparison, *Science* has around 1,000 citations, *New Scientist* more than 2,300 and *The Times* an impressive 19,603. The new electronic version of the dictionary is published by Oxford University Press on a single compact disc priced at just under £600.

Peter Tallack

House vote to kill SSC reflects squeeze that budget deficit puts on all research

Washington. Last week's vote by the US House of Representatives to kill the Superconducting Super Collider (SSC) surprised both supporters and opponents of the \$8,500-million project. As a result, neither politicians nor scientists know whether the vote is a temporary setback or a fatal blow to the proton-antiproton accelerator being built in Texas.

Nevertheless, the impact of the 232:181 vote to eliminate all but \$34 million of the \$650 million requested for the fiscal year beginning on 1 October is being felt far beyond the high-energy physics community. Some scientists believe that it signifies a lack of support for any costly basic research project that does not contribute relatively quickly to public health or economic well-being. Others say that the vote was merely the first opportunity for Congress to show that it is serious about reducing the \$400,000-million deficit since members defeated a proposed amendment to the US constitution two weeks ago that would have required them to adopt a balanced budget each year.

What everyone agrees upon is that Congress is almost certain to reduce the sizeable increases for research proposed by the president, George Bush, in his 1993 budget. In

particular, officials at the National Science Foundation (NSF) have abandoned all hope for an 18 per cent increase, the US National Aeronautics and Space Administration (NASA) fears for the future of its space station Freedom and the US National Institutes of Health (NIH) would be pleased with a budget that allows them to keep pace with inflation.

The source of this pessimism is a 1990 agreement that limited annual increases in federal spending for the next five years. Research has so far managed to escape the worse effects of that agreement, in part because of the administration's strong support for most areas of basic science and in part because Congress has given to basic science a good portion of 'new' money it found over the past two years.

That money no longer exists, and the relevant category for science (known as domestic discretionary spending) is being squeezed by other programmes with larger constituencies or greater political clout. For example, Bush had proposed nearly \$1,000 million in increased spending for NASA and NSF next year. But the subcommittee that funds the two agencies also allocates money for veterans' affairs, for which Bush is also seeking another \$1,000 million. When

the subcommittee looked at its yearly allotment, it found that it had money for only half the increase that the president wanted. Forced to choose, the members took the political route; the increase in science programmes would be cut to a quarter of the president's request, while medical care for veterans would receive two-thirds of its increase. That distribution could well be changed in the weeks to come, however, as the bill moves through both houses of Congress.

Because Congress has so little room to manoeuvre, this year's federal budget is much more likely than in the past to resemble the president's initial request submitted in January. That poses a special problem for NIH. In recent years, legislators have added liberally to the president's requested increase. But this year NIH must compete with equally popular pre-college education programmes for a pot that is not big enough for full increases in one, much less both. Congressional staff members say that NIH will be lucky to receive even the president's request — an increase of about \$440 million, or five per cent, which matches the rate of inflation in the cost of biomedical goods and services.

With respect to the SSC, there are two versions of what is likely to happen next. In the first, the Senate would also vote to eliminate funding for the SSC and the project would be dead. In the second, the House would be persuaded to change its mind in conference with the Senate after the Senate voted sufficient funds to keep the project alive. A subcommittee of the House two weeks earlier voted to hold the SSC to its current level of \$484 million, and that figure is believed to be the largest possible amount that could emerge from the Congress.

In the meantime, the vote demonstrates that the prospects for the SSC are far gloomier than had been thought. Last year, an attempt to kill the project failed by 87 votes, and neither side expected that margin to change very much in last week's show-down vote. Opponents took credit for having cast doubt, in a series of hearings, on the government's ability to manage the project properly and hold down costs, and they feel that the tide has run out on the SSC.

"Even if the Senate succeeds in saving it this year", says Dan Pearson, an aide to one of the collider's harshest critics, US Representative Sherwood Boehlert (Republican, New York), "there will be a lot of fresh faces in Congress next year who have no track record [of support for the project] to defend and who have promised their constituents that they will balance the budget. I don't see how it can make a comeback."

Jeffrey Mervis & Christopher Anderson

Japan 'finds' money for research

Tokyo. Pressure by a group of Japanese politicians to increase government funding for science appears to have paid off. In budget ceilings for next fiscal year (starting April 1993) set last week by the Ministry of Finance, a new category of extra funds has been introduced that is expected to help end a steady deterioration in the infrastructure of Japan's government laboratories, in particular the universities, over the past decade.

The amount going to science will be determined in budget negotiations between agencies over the next several months. However, government officials expect the new category to remain in the budget beyond 1993, bringing much needed long-term relief to the country's research facilities.

A science pressure group headed by Kishiro Nakamura of the ruling Liberal Democratic Party (LDP) has been pressing the finance ministry over the past few weeks to commit more money to science. The decision follows a series of recommendations from various government committees and agencies to increase the government's science budget, as well as a decision by the cabinet in April to double government spending on research and development "as early as possible".

At present, under a policy of severe fiscal restraint that has been in operation since the early 1980s, the government budget for general operating expenses in national institutes and universities must decrease by 10 per cent a year. The new fund will add 5 per cent, or about ¥215,000 million (\$1,700 million), to the total running expense account for all government ministries and agencies. Part of the funds are earmarked for 'academic research'. A particular beneficiary is expected to be the Ministry of Education, Science and Culture, which runs Japan's universities. The Agency of Industrial Science and Technology of the Ministry of International Trade and Industry, which has several national research institutes, is also expected to apply for some of the extra funds. One agency not expected to share in the largesse is the Science and Technology Agency, which does not operate its own facilities and which receives most of its money through other budget categories.

David Swinbanks

Clinical research damaged by UK health service reforms

London. Clinical researchers in London's teaching hospitals say that the rapid pace of reforms to the National Health Service (NHS) are damaging the quality of their work. Uncertainties over funding are making them unwilling to initiate new projects, while changes in the patient referral patterns are posing obstacles to existing work. One professor at University College Hospital called the changes an "insidious erosion of clinical research".

The Department of Health (DoH), which is reviewing research funded by the NHS, says that the only evidence for their position is anecdotal. Yet many researchers say that the damage is hard to detect because the work straddles the border between clinical practice and basic research. In fact, DoH hopes that its review will make clear what type of research the NHS is funding, and who is doing it.

The reforms were introduced in April 1991 to improve the NHS's role as a provider of primary health care and to provide health authorities with greater financial independence. One result has been increased pressure to send patients to hospitals that offer the best value for money, which translates into a preference for treating patients in their home district.

This preference has hurt London's 14 teaching hospitals. The capacity of London hospitals far exceeds the health care needs of the local population, prompting repeated calls for reductions. In addition, the staff at these hospitals serve as teachers, clinical researchers and practitioners. At the same time, the division between their academic and clinical roles can be blurred. Staff employed by the NHS to perform routine care often become involved with research, and some academics contribute to clinical care.

All the teaching hospitals have suffered to some degree. The most recent sign of stress is the announcement by Bloomsbury and Islington Health Authority, facing a deficit of £20 million, that it will close 100 beds and eliminate 200 staff at University College and Middlesex Hospitals. Together, these two hospitals make up London's largest medical teaching complex.

Because the number of beds determines the number of students, and the funding tends to follow the students, fewer beds means less money flowing into the teaching hospitals. Most teaching of medical students involves a research component, and decreased funding also threatens to weaken research groups.

Researchers also fear that the cost-conscious health authorities will not enrol patients in clinical trials that involve expensive developmental drugs and will refuse to

send them to hospitals where new techniques for diagnosis and therapy are being investigated. Those restrictions would erode their patient base, which constitutes a valuable resource. "These projects are often the research that solves immediate clinical problems, or that form the seed corn for longer term fundamental academic research," says David Delpy, a professor in medical physics at University College, London (UCL).

Cost considerations have already led health authorities to prohibit some doctors from referring patients to leukaemia trials. A report* from the Universities Funding Council (UFC) on the effects of the reforms expresses concern for the research taking place in specialist clinics for common disorders such as diabetes, hypertension and asthma. In the long term, there are fears that the NHS may devote most of its resources to common conditions and abandon work that does not have an immediate financial benefit.

The NHS has been undergoing a review of its research spending since it launched its first research strategy just over a year ago (see *Nature* 351, 7; 1991), with a pledge that research spending would be increased over the next four years. At present, around £250 million (US\$440 million) is spent a year,



Closed wards reflect overcapacity problem.

representing about 1 per cent of the NHS budget. The aim is to reach 1.5 per cent by 1997–98.

However, until the review is complete — and until the NHS completes its analysis of where it is currently spending money on research — there is considerable concern over which areas of research will be favoured. "There is a danger that there will be a reluctance to plan academic ventures when the future is so uncertain," says David Linch, professor of haematology at UCL. The first results of the review were visible last week, when the central research and development committee of the NHS sent out an announcement soliciting proposals for research projects in certain areas of mental health.

Ian Mundell

* Second Report on the Effects of the NHS Reforms on Medical and Dental Education and Research (Universities Funding Council, Northavon House, Coldharbour Lane, Bristol BS16 1QD, UK.)

Russian academy vote excludes Jews

Moscow. There is consternation in the Russian Academy of Sciences (RAS), and outside as well, after this month's election of new members: almost all candidates with Jewish surnames were blackballed. Many academicians themselves are dismayed by this demonstration of antisemitism among the scientific elite.

The outcome was unexpected. Usually, the general meeting of the RAS formally approves choices made earlier by its 18 academic divisions, subject only marginally to the settling of old scores. But this month, nearly all the candidates proposed by the departments of economics, international economics, philosophy and law and nuclear physics were rejected.

The list of rejections from nuclear physics includes such prominent figures as Semyon Gerstein, Isay Gurevich and Ilya Feinberg, all with Jewish names. But Alexander Mikhailov was also rejected, apparently because of the misfortune of having recently become a government minister.

In the event, the elections were extraordinary. Some successful candidates would not have been elected conventionally, including members of the new Nomenclatura

such as Vladimir Shorin (chairman of the Parliamentary Commission on Science and Technology), who becomes an academician, and Ruslan Khasbulatov, the Speaker of the Russian Parliament (corresponding member). Also elected academician was Igor Shafarevich, an eminent mathematician who is also a well-known ideologist of antisemitism.

Indignation at the results, and the conviction that they had been unfairly overlooked, then prompted the corresponding members of the RAS to demand that they should automatically be promoted to full membership. The leadership promised instead that there would be a further round of elections.

After the elections, the president of the academy voiced his surprise at the outcome, but Academician L. I. Abalkin may have been the only one to speak openly of antisemitism. Abalkin criticized the election system for allowing an expert in the humanities or a microbiologist to vote on the candidacy of a physicist or a jurist. But the results of this month's election suggests that the trouble does not lie with the electoral system alone. **Vladimir Pokrovsky**

FBI attaches strings to its DNA database

Washington. Researchers are criticizing the US Federal Bureau of Investigation (FBI), which controls one of the largest databases of human DNA samples, for refusing to allow an independent researcher to use the data unless FBI scientists are permitted to review his work and to be listed as coauthors. Although the FBI says that the ruling is not agency policy, officials admit that they have not yet decided how to handle the issue of data access. A recent report by the US National Academy of Science (NAS) strongly recommends that data be made freely available to all researchers.

In a letter dated 12 May to Seymour Geisser, a University of Minnesota statistician, James Kearney, the head of the FBI forensic science research centre, refused to permit the use of the data unless Geisser agreed to give an FBI scientist the right to approve the paper. Kearney says that the FBI is "not quite sure of [Geisser's] intent" in seeking to analyse the data, pointing out that Geisser has testified for the defence in DNA fingerprinting cases. Kearney acknowledged that the FBI has provided the data to other researchers (he named three scientists, at least two of whom have testified for the prosecution in DNA fingerprinting cases), but said they were picked because they were experts in population genetics.

Kearney agreed to provide the data only on the condition that Geisser analyse them in collaboration with Bruce Budowle, an FBI scientist, and that Budowle should have final approval of the paper. Geisser rejected that arrangement as "obviously unacceptable". He has since revised his paper — on statistical techniques as they apply to population data — to use an artificial dataset.

Geisser had asked the FBI for permission to use the data on the advice of Charles Epstein, editor of the *American Journal of Human Genetics*, which is considering Geisser's paper. Although the data had been released as part of a court case, Epstein recommended that Geisser seek explicit approval from the FBI for their inclusion in the paper. But he decided to drop the FBI data entirely after the FBI refused to grant permission and after Geisser said he was unwilling to let the agency researchers "censor" the paper.

"It's an unfortunate situation when such data are not made available, presumably because they [the FBI] might not be in agreement with the outcome", says Epstein. "By trying to extract that kind of promise [of coauthorship] they are showing that they are not truly interested in the free flow of information."

Earlier this year, the same paper triggered a dispute with the FBI after federal prosecutors demanded that Geisser turn over a draft of the manuscript and any comments

from reviewers. Because this request came 15 minutes after Epstein faxed the comments to Geisser, the case prompted allegations of conspiracy (see *Nature* 355, 753; 1992), although most of the parties now say that the timing was an unfortunate coincidence.

The data have been used by the FBI in several court cases to support convictions based on DNA fingerprinting and have been the basis of published articles by several pro-DNA-fingerprinting researchers. According to Victor McKusick, chairman of the panel that produced the NAS report, the data are generally considered to be in the public domain.

McKusick says the FBI position of re-

"We are willing to approve your use of FBI population data with certain provisions....The FBI data may be used only in a joint collaboration with Dr Budowle....The use of the data is restricted to this one paper. All parties (i.e., authors) must agree to the entire contents of a final manuscript prior to submission to a journal. Any changes whatsoever in the manuscript must be agreed upon by all collaborating parties."

— FBI letter to Seymour Geisser

quiring coauthorship on Geisser's paper "is contrary to the spirit" of the NAS report, which explicitly calls for open access to population data. Noting that "presenting scientific conclusions in a criminal court is at least as serious as presenting scientific conclusions in an academic paper", the report concludes that "the data underlying [those conclusions] must be freely available....If scientific evidence is not yet ready for both scientific scrutiny and re-evaluation by others, it is not ready for court."

John Hicks, assistant director of the FBI's laboratory division, says that the letter from Kearney does not restrict access to the data. Although the FBI specifically warned Geisser that use of the data was contingent on FBI coauthorship, he says, "we can't stop Geisser from using them without our permission." It would be a "misinterpretation" to read the letter as a prohibition on use without complying with the FBI demands, he says; Kearney was "simply asking that a professional courtesy [coauthorship in exchange for use of data] be extended".

Hicks says that FBI policy on data access is "still evolving". The agency is following a recommendation from the NAS report to form an advisory committee to recommend

data policies. The issue, however, is an "awkward" one, Hicks says — "privacy experts are concerned about people [using the data to] draw possibly adverse inferences" about the behaviour of population subgroups. "In general we agree with open and free access, but I'm a little bit concerned that if we do it, someone may misuse it."

But McKusick calls that argument a "smokescreen" to restrict use of the data. Because the particular DNA regions used in the database do not correspond with expressed traits, "nothing in the data identifies whether the individual is a rapist or a Supreme Court justice". Concern about privacy need not be a barrier to the free and open use of the data, he says.

Christopher Anderson

Germany plans university reform

Munich. Responding to a call for reform of the university system in Germany, the Ministry of Education and Science last week announced a six-point plan to speed students through their courses and improve university facilities, particularly for scientists.

Last month, the regional governments of Germany's 16 states spoke out for the first time in support of their financially strapped universities (see *Nature* 357, 349; 1992). In response, the education minister, Rainer Ortleb, has recommended that:

■ Universities should require students to graduate within four years instead of the current average of eight years.

■ More places should be provided at universities specializing in applied science research. Ortleb hopes to increase the proportion of students attending these vocational universities from 28 to 40 per cent.

■ The government should invest DM1,600–DM2,000 million (US\$1,000–\$1,300 million) next year in building projects, and vocational universities should be given priority.

■ More science graduates should be encouraged to enter the university system to be trained to replace the steadily ageing population of professors, 50 per cent of whom will have reached retirement age by 2005. Ortleb has recommended that spending on university research should remain constant in the next few years, despite Germany's recession, to give scientists the opportunity for suitable training.

The federal government will discuss the recommendations and their budgetary implications this summer. A final decision, expected by the end of the year, will also depend on approval of the universities principals and trade unions. Allison Abbott

Without a way to measure their success, Japanese projects are very hard to stop

Tokyo. The lack of an independent system to monitor and assess research has made it extremely difficult for Japan to end large national research projects once they are under way. The decision by the Ministry of International Trade and Industry (MITI) to keep alive the fifth-generation computer project (see accompanying story) after a half-hearted review is just one example of a government project that continues to consume public funds despite serious shortcomings.

"In Japan, national projects are very hard to start, but once they get going they are almost impossible to stop", says Robert Geller, a geophysicist at Tokyo University and a vocal critic of Japan's national earthquake prediction programme, which has been running with little change for nearly 30 years. Assessments of big government projects tend to be carried out by insiders who, Geller says, "are very positive".

The fifth-generation computer project is no exception. A committee of 27 academics and industrialists, headed by Hidehiko Tanaka of Tokyo University, were asked by MITI to review the project after its initial ten-year life ended in March. Nearly all the committee members have close associations with the project, including executive director, Hiroichi Hiroshige, and the head of the project's research institute, Kazuhiro Fuchi.

Predictably, the report released last week concludes that the "original goals of the project have been achieved". It points out that 70 per cent of overseas researchers rate the project as "superb or excellent".

In reality, the fifth-generation project has fallen far short of its repeatedly stated goal of making a "user-friendly" parallel computer with 1,000 processors. The largest machines on display at a project conference earlier this month contained only 256 processors, and US researchers with access to them say the machines are very hard to operate because the programs they run on can be written and understood by only a very limited group of experts. MITI's proposal to extend the project and make the computer compatible with standard computers (see below) is a clear indication of its present limited use.

One of the MITI committee members admits that the assessment is "just decoration intended to praise the project". He says that the committee meetings were very short and the agenda set by MITI and leaders of the project. Some eminent committee members did not attend a single meeting. The general opinion of computer scientists in Japan is more negative, he says; although the project has stimulated the research com-

munity, it clearly fell short of its "overambitious technical goals".

A similar criticism can be made of Japan's earthquake prediction programme. The programme's leaders have argued that earthquake prediction is a worthwhile goal; some even claim that the goal has been reached. Yet most of the rest of the world believes that earthquake prediction is beyond the present capabilities of science, and has switched its attention to hazard mitigation.

In the United States and other Western countries there are independent organizations that review and monitor major government projects. For example, the US Office of Technology Assessment (OTA) was created in 1972 to advise Congress, and such organizations as the US National Academy of Sciences (NAS) also play this role. But Japan has no equivalent organization.

The Science Council of Japan, an advisory body to the government composed of several hundred prominent academics, is the nearest counterpart to the NAS. But the council fell out of favour with the government in the 1970s because of its constant criticism of government policy and the belief of the ruling Liberal Democratic

Party (LDP) that the organization was run by communists.

In 1985, the system of election of council members was restructured to allow the prime minister to screen all candidates, and relations with the LDP have since improved. But the council's budget is so small that it cannot even afford to pay travel expenses for its members.

Jiro Kondo, president of the council, made a proposal backed by the Japan Society of Technology that the council should take on a role equivalent to the OTA (see *Nature* 348, 572; 1990). Kondo pointed to the nuclear-powered ship *Mutsu*, which has spent only a few days at sea despite costing ¥120,000 million (nearly US\$1,000 million), as an example of a government project that should be terminated.

In the meantime, *Mutsu* continues to receive thousands of millions of yen a year while sitting in port. The Science and Technology Agency is planning to rip out the ship's nuclear engine, replace it with a diesel one, and convert the ship into an oceanographic research vessel. But oceanographers point out that the ship was never designed for oceanographic purposes.

David Swinbanks

Fifth generation project lives on

Tokyo. After spending more than ten years and ¥50,000 million (US\$400 million) to develop a fifth-generation computer, a machine supposed to revolutionize the world of computing, Japan's Ministry of International Trade and Industry (MITI) plans to spend another ¥4,000 million and two years to make the computer compatible with commercially available machines. MITI's decision, announced last week, represents the final ironic twist in a project that, when announced in the early 1980s, sparked fears in the West that Japan would capture the world's computer markets with a new generation of machines that would make conventional computers obsolete.

MITI and project officials justify their decision to extend the project, which has already lasted one year longer than planned (see *Nature* 356, 273; 1992), by pointing to a glowing assessment by academics and industrialists closely associated with the project (see accompanying story). But the new effort, devoted to linking the Japanese computers to those that use the US-developed UNIX operating system, serves instead to highlight the shortcomings of the original project.

In the past two years, the project's Institute of New Generation Computer Technology (ICOT) has donated terminals to the US Argonne National Laboratory, the US National Institutes of Health and the Lawrence Berkeley National Laboratory. ICOT publicized the donations as a joint US-Japanese effort to develop fifth-generation software. But US researchers now see ICOT's move as largely a publicity stunt, and complain that the fifth-generation computers are slow, cryptic, prone to errors and incompatible with any other computer.

Shunichi Uchida of ICOT acknowledges the complaints. But he and MITI officials say that the aim of the extension is to overcome such problems and make the fifth-generation computer available to everybody. Uchida, however, notes that the speed of the fifth-generation computer will be reduced to "half" in linking up to UNIX computers.

The extension plan requires the approval of the Ministry of Finance. But the finance ministry seldom rejects such proposals, and there is no system for external technical review.

David Swinbanks

Research effort is preserved in plan to close London Zoo

London. Research at London Zoo will continue unaffected by the decision to close the Regent's Park site to the public at the end of September (see Commentary on page 621). All the same, research staff are supporting efforts by their colleagues to put together a survival package based on the zoo's role as a conservation centre.

The situation has been carried over from last year's crisis (see *Nature* 350, 453; 1991), which arose when the Zoological Society of London, which manages the zoo, announced that reductions in staff and services and additional government support were needed to keep the zoo open. Although the government did not provide any more money, the publicity afforded by the crisis increased attendance sufficiently to allow management to abandon its original plan. A 'break-even' budget was implemented in March, but it has proved difficult to sustain. The zoo now faces an annual deficit of £2 million (US\$3.5 million), and the society says that it will have to close the zoo in order to maintain its other activities.

The zoo's research, carried out at the Institute of Zoology, will be affected very little by the closure. The work does not depend on the animals at the Regent's Park site, and relies instead on field work and cooperative efforts with other zoos and wildlife parks. However, the staff at the institute have strongly supported the London Zoo Survival Group, which includes some 80 per cent of the zoo's 240 employees. The group's conservation rescue plan would allow the researchers to maintain a close link with an active animal collection.

The institute is largely immune from the zoo's financial crisis. It receives £1.3 million a year from the Universities Funding Council, via the University of London, and has grant income of £827,000 from sources such as the research councils and the Royal Society for the Prevention of Cruelty to Animals.

The activities that the society says it is protecting by closing the zoo include conservation projects overseas, in other zoos and at its other site, the Whipsnade Wild Animal Park. However, zoo staff claim that much conservation work that is taking place as part of the curatorial process will be lost if the animals leave. Of particular interest is the insect house, where a number of large, genetically diverse populations of endangered species are being preserved for re-introduction into protected habitats. These include Mexican red-kneed bird-eating spiders and wart biters, a rare species of giant British cricket.

Gaining control of the zoo will be the crucial step for the survival group. With

representatives from all sections of the zoo's staff the group also claims a sizeable following among the fellows of the Zoological Society, with whom the decision ultimately rests.

Chief among the group's complaints is that the zoo management has failed to publicize the zoo's activities as a conservation body. By advertising a trip to the zoo as a "good day out", says Doug Richardson of the group, the zoo is out of step with current public opinion towards animals in captivity. The survival group feels that an appropriate publicity campaign would draw the necessary financial support from public and private sources.

However, the most prominent external package to rescue the zoo would turn it even more into a theme park. New Zoo Developments (NZD), a consortium consisting of the Laing construction company, bankers Samuel Montagu and US designers Cambridge Seven Associates, has proposed a £61 million scheme which would include a rainforest pavilion, a coral-reef aquarium with an underwater tunnel for visitors and a number of other naturalistic settings for the animals.

Ian Mundell

Lab death blamed on gas build-up

Washington. The explosion of a 'cold fusion' cell at the Stanford Research Institute (SRI) in California earlier this year, which killed researcher Andrew Riley (see *Nature* 355, 102; 1992), has been attributed to an unnoticed build-up of oxygen and deuterium inside the cell. That conclusion, based on a five-month investigation, dispels speculation that the explosion might in some way have been related to excess heat generated by the phenomenon itself.

The explosion occurred in a closed calorimetric cell, in which oxygen and deuterium liberated by the electrochemical reaction were normally recombined into water by a platinum catalyst. An expert team assembled by SRI concluded that the catalyst failed after 800 hours of running, allowing the gases to accumulate, and that the build-up of pressure forced a piece of teflon into a wire-hole that would otherwise have allowed the gases to escape.

The precise cause of the explosion could not be pinned down, but the report speculates that when Riley moved the cell, part of the platinum catalyst was re-exposed and began to work again, growing hot and eventually igniting the gas mixture.

David Lindley

New Zealand tries to bring science closer to industry

Sydney. The New Zealand government is overhauling the way in which science is organized in an effort to encourage industry to invest more heavily in research and to increase links between government scientists and the private sector.

On 1 July, the Department of Scientific and Industrial Research (DSIR), the largest government research organization, will combine its research operations with those of other government departments to form ten private companies wholly owned by the government. The companies, to be called Crown Research Institutes, will employ about 3,000 scientists. About 70 researchers will lose their jobs in the move, which involves the science divisions of the Ministries of Forestry and of Agriculture and Fisheries, as well as the Meteorological Service and the DSIR.

The reorganization will free government scientists from cumbersome procedures and severe restrictions on their commercial activities that included a prohibition on receiving royalties from their inventions. The new structure will allow individual institutes to carry out such activities as contract research, joint ventures and applied research with industrial partners.

Apart from making it easier for government scientists to attract private funds and to accept contract work, the new arrangement is intended to improve outside access to government research facilities. This access, the government hopes, will also inspire industry to increase its present low level of spending on research and development.

The changes will also affect the way in which government scientists are funded. Instead of giving money directly to the institutes in a block grant, the government will require scientists to submit proposals to a separate set of bureaucrats in the Ministry of Science. The money will be allocated in grants from a Public Good Science Fund. The fund will contain all but NZ\$185 million (US\$100 million) of the NZ\$435 million that the government spends on science. About NZ\$110 million is given to university faculty for basic research, and the rest is spent on medical and military research.

The new arrangement, a reflection of the ruling National Party's belief that public servants who control funds must be separated from those who spend the money, will force institutes to compete against one another for funding. A committee will select the best proposals.

The reorganization will not increase government spending on research. But it is expected to stabilize for several years the level of funding after a decade of declining support for science.

Mark Lawson

The zoo that has to close

Lord Zuckerman

London Zoo is the only major national zoo run without the benefit of substantial support from public funds. No wonder it is having to close.

THE impending closure of the Regent's Park Zoo is immensely saddening, but comes as little surprise to those who have followed its fortunes over the past few years. The present council of the Zoological Society of London (ZSL) is an educational charity, which until 1964 was entirely supported by gate-money, by fellows' and members' subscriptions, and by the occasional benefaction. The bulk of all its resources up to 1932 went to the upkeep of the London Zoo, and after that date to Whipsnade Park as well. What was left covered the normal operations of a scientific society — meetings, library and publications.

When I became the society's honorary secretary in 1955, attendances fluctuated around the 2 million mark, with gate-money sufficient to cover the menageries' running costs. What reserves had accumulated during the Second World War and first postwar years were used to build an animal hospital, to establish an assured pension fund and to cover the costs of a supplies and workshop building. Funds were provided by the Nuffield, Wellcome, Ford and Wolfson foundations for the building of research laboratories, and for modernizing the library and meeting room. A few wealthy individuals donated the money to build new animal houses, shortfalls being covered by bank loans.

The first government money the society ever received was a grant in 1964 of £250,000 plus a matching long-term loan. When a further loan of £375,000 was arranged five years later, the government appointed professional consultants to see if the society could improve its financial management. Their conclusion was that the only way to complete the zoo's rebuilding programme would be for the government to help. Repayment of the 1964 loan was then waived, a grant of £650,000 provided to repay other short-term debts and one of £700,000 to help the society over the period 1970–76.

By the early 1980s, it was obvious that the society could no longer continue to run a major public amenity. In the 1950s, salaries and wages accounted for less than half the society's total expenditure. By 1980, the proportion had risen to two-thirds, exclusive of contributions to the pension fund. Gate charges were being raised to keep pace with increasing staff costs, and staff were being laid off.

Fortunately the society's research staff were supported mainly by the research councils on the basis of merit.

The financial position deteriorated acutely between 1980 and 1983, my final years as an officer of the society. But my message to the council and Whitehall remained the same — the society was the custodian of the only national zoo unsupported by public funds. If the London Zoo were not to close, the government would have to keep it open.

As negotiations continued, help was provided on an annual basis with new government-appointed management consultants recommending that an 'operational plan' should be prepared to show how the society could become solvent, but with the London Zoo still under its wing. The upshot was that by the time I retired from the presidency, the government had committed itself to providing the society with up to £2 million a year for the two years to March 1986.

Without any other money in sight, a management committee was formed and set about the promotion of the zoo's image by improving its public amenities. Little improvement in attendances resulted, and in 1988, in response to pressure from Lord Peyton, then the society's treasurer, and, on the recommendation of another firm of management consultants, the government provided a 'once-for-all' grant of £10 million for the zoo on the condition that a subsidiary company, Zoo Operations Ltd, should be set up to manage the menageries. Commercial and advertising executives were recruited to succeed where the council had presumably failed. Considerable publicity followed, but little else. No new animal houses were built. The company spent money freely, and business partners were vainly sought for the task of running the zoo as a commercial venture.

Little more than a year ago, Sir Barry Cross, the society's secretary, issued a statement saying that over the past eight years the society had had little success in raising money from private sources, and that only £4.5 million remained of the government's final £10-million grant. Operating costs were running at £6.3 million — well above what could be afforded. Were the zoo to close, the society also faced a dilapidations bill of £13 million.

Later that year, several members of the council, including the treasurer, res-

igned *en bloc* — without publicizing the reason. With attendances continuing to fall well below the million mark, and with no money in sight to cover its labour-intensive operations, the new council had no option but closure.

As landlord, the Crown Commissioners can now take over without compensation several listed and famous zoo buildings. It is conceivable that some may continue to house a few animals — it is sad to think of Lubetkin's famous pool without penguins. The bulk of the society's live collection will be kept at Whipsnade, little further from the centre of London than the zoo was when it started in 1826. The ZSL's main building, library, meeting rooms and laboratories will continue where they are.

Without the incubus of the zoo, the society is financially viable when its accounts are viewed in a way appropriate to a scientific institution, and not, as now, as though it were a commercial undertaking. For example, in the 1990–91 report, the library is shown as running at a deficit — which is ridiculous. When I enquired, I was told that the deficit represented the library's share of the society's overall shortfall of £4.6 million. On the other hand, no part of the deficit is shown as attributable to the finance department. What the accountants of Zoo Operations Ltd clearly do not realize is that if the ZSL were not a charitable scientific body that publishes and possesses a library, it would never have been allowed 30 acres of public park rent-free. What rent a commercial menagerie would have to pay, I cannot bear to think. The menagerie element of the society also benefits from the fact that, as a charitable scientific institution the society pays less than half the local taxes that the local council would otherwise demand.

Because of its charitable status, the society has for four years indirectly helped to finance the menageries. Having to close the Regent's Park Zoo is highly regrettable, but Field Marshal Sir John Chapple, the new president, and his council are nonetheless to be congratulated on having had the courage to show that not Zoo Operations Ltd but the council remains master of a world-renowned scientific institution.

Lord Zuckerman is emeritus professor at the University of East Anglia, Norwich NE4 7TJ, UK.

Virtues of small universities

SIR — The continuing shortage of funds for research in science (and engineering) and the undoubted need to be selective in the distribution of those funds that are available has raised again the idea of a super-league of 'research universities', establishments that are alone able to pursue research across a wide range of subjects.

But I should like to mention some of the hazards, particularly in view of the impending transition of the polytechnics to university status and the transfer to research councils of some of the funds for research in universities previously handled by the Universities Funding Council (more than £100 million per annum). These changes will certainly increase the competition for research funds.

Of the research councils, the Science and Engineering Council has by far the biggest budget and its chairman, Sir Mark Richmond, is a strong proponent of the idea of "a cohort of research universities" (see, for example, *Nature* 353, 379; 1991).

Although in some branches of some subjects very large groups and infrastructures are required and these can perhaps best be provided by large universities, most research is not in this category and small groups (even single individuals) are more the norm. Such groups can flourish in universities of very different sizes.

A relevant and very obvious reason for distributing much of our research effort is the importance of at least some of each undergraduate's teachers being exposed to research. Although all researchers are by no means good teachers, the enthusiasm shown by a committed researcher is invariably contagious, not to mention the value of teachers of advanced courses being up to date.

Even in the areas of big science where big supporting facilities are required, these are often being provided outside individual universities. No doubt this trend will continue. A prominent example is in astronomy, where the necessary large telescopes are provided nationally, or more often internationally, on remote mountain tops. Research groups in small universities are just as likely to succeed with their research projects using these telescopes as those from larger institutions. Peer review of applications for observing time, or for the construction of the necessary sophisticated instrumentation, ensures that quality is maintained.

A crucial point of growing importance in scientific research in general relates to the analysis and interpretation of ex-

perimental data and to the related area of 'theory'. To achieve a sensible scientific return from the effort devoted to the collection of data from the big 'machines', whether accelerators, synchrotron sources, telescopes, satellites or whatever, wide dissemination is vital. Ease of communication means that even small groups need not feel isolated. Indeed, small groups may be superior in this area to those whose resources are inevitably stretched to provide the hardware. A balance is surely called for.

Another reason for setting our face against over-concentration of research in a small number of universities is the resulting stagnation that would occur. 'Selectivity' by all means, but let us have competition — group against group. The virtues of large universities, where they exist, will shine through automatically in their groups having the stronger case for support.

What, in fact, are the perceived advantages of the large universities and how valid are they?

(1) Large universities have better facilities in the form of workshops, computing, libraries and so on. This is partly true but the increasing specialization of workshops (which means that small establishments with important research groups can have matchless facilities), the increasing general availability of powerful computing networks and the increasing use of easily accessible databases nullify many of the advantages.

(2) Large universities have stronger interdepartmental connections. Administrators like to think this but the facts are often otherwise. Interdepartmental rivalries often militate against really effective collaboration. The efficiency of Interdisciplinary Research Centres where the constituent departments are drawn from different (not large) universities shows that associations of small institutions can be very effective.

(3) Large universities can offer 'economies of scale', the most important of which probably relates to the question of the competition for an academic researcher's time. In big universities, unless there has been undue fragmentation of undergraduate courses, big departments have staff who spend less time on non-research activities. In small universities, or more particularly, in small departments, teaching and administrative duties can make excessive demands on an academic's time. The remedy is straightforward: small universities with research aspirations should endeavour to concentrate their teaching on a limited number of well-structured courses. A plethora of 'options' should be eschewed. With undergraduates being

educated over a wide base, on entry, but to less depth than hitherto, fewer options rather than more should be the order of the day.

The requirements for a healthy research effort in a small university appear to be the following:

- (i) Choice of research topics to fit the criteria of departmental and group size;
- (ii) Concentration of resources on departments/groups that demonstrate their ability and suitability;
- (iii) A realistic attitude to the need to secure adequate time for research.

If these conditions are met, I can see no reason why a small university should not have almost as big a fraction of its staff actively engaged in front-rank research as any large university.

A. W. Wolfendale

(Astronomer Royal)

Department of Physics,

The University,

Science Laboratories,

South Road, Durham DH1 3LE, UK

Text readability

SIR — What Donald Hayes says (*Nature* 356, 739; 1992) about the relative readability of texts is fallacious. The purpose of language is to convey meaning and this is not related to the number of words used but to the precision with which the symbols convey meaning.

The definition of a mathematical concept such as a vector space is the same in all languages and is extremely precise. It does not have the multivalent meanings attached to words. Once the barrier to entry is passed, the mathematical expression is very much simpler than the confused discourse of philosophers and journalists.

In the *Times Literary Supplement* of 8 May 1992, page 12, the mathematical term 'parameter', to which mathematicians assign a one-valued meaning, was used by a cabinet secretary in several impenetrable different senses, possibly as meaning 'limits', 'criteria' or 'considerations'. In the medical literature, the word is used for 'variables', 'constants', 'conditions' or for 'just damned fate'.

Another factor in readability of texts is what one inserts from outside sources. The specialization of knowledge makes reading a cryptic art for those outside. In other words, readability cannot be measured because it is a function of transference from other sources than the particular text. Obvious examples occur in law. The opinions of the judges are replete with reference such as: "the *ratio decidendi* of Dickson C. J. in *R. v. Martineau*."

Daniel L. Turner

22 East Sedgwick Street,

Philadelphia, Pennsylvania, 19119, USA

Optimistic, independent, aggressive: Texas shoots for the top in science

TEXAS is a state of mind. It is optimistic, bold and independent. When we think of Texas, our imaginations conjure cowboys and cattle ranches and black gold gushing from the Earth. We imagine a state populated by the likes of J.R. Ewing, the Dallas oil baron who made his home on the Southfork Ranch. It is easy to forget that J.R. is only a figment of a television writer's imagination, especially when the real Southfork, complete with a museum of memorabilia from the television series, is really for sale to the highest bidder.

Texas is space — vast flat prairies of sagebrush and tumbleweed and wild flowers. It is also outer space, locked into our memories as the ground-based centre of mankind's journey to the Moon.

Texas's scientific institutions, even by US measures, are young. Many are making a determined effort to achieve world-class intellectual status, often by attracting scientific stars with generous financial backing from local business tycoons. First among them is H. Ross Perot, the self-made billionaire whose undeclared and unconventional determination to become the next president of the United States fits perfectly with our image of the Lone Star State.

Their goal is to make Texas the undisputed scientific capital of the United States. It is a lofty ambition. But then, it is no surprise that Texans want Texas to be great.

During the past several years, *Nature* has surveyed science and technology in a number of nations in its semi-annual 'country' issues. This time, the journal decided to take a look at Texas, a state whose

physical size — it is possible to travel 800 miles in a straight line without leaving the state — and population — 17 million — make it larger than many countries. We found a state that, like the rest of the United States, is financially beleaguered. The oil glut of the mid-1980s lowered oil prices and forced Texans to admit, for the first time in decades, that even they are not made of money.

But Texas also appears to be a state whose leaders (most certainly its scientific

leaders) are an optimistic lot. Visit the principal science campuses of California or New England and you will hear that the research enterprise is in jeopardy, being strangled by insufficient funds, too many regulations, and state and national legislators who fail to appreciate the importance of science.

Visit Texas and you hear stories about institutions on the verge of substantial growth. The University of Texas South-

western Medical Center at Dallas, for instance, is constructing six new state-of-the-art research buildings on 30 acres donated by the John D. and Catherine T. MacArthur Foundation. In San Antonio an entirely new research park is being created on 1,500 acres

of prairie outside of town. Texas researchers like to say that their academic institutions, in contrast to their academic counterparts in the East, have ample room to expand.

Houston, where two dozen new biotechnology companies have been founded since 1983, is anxious to become the hub of the state's efforts to compete with other biotechnology enterprises. Texas A&M University in College Station, Texas's leading agricultural research institution, 100 miles northwest of Houston, is among those that look to the future and see biotechnology. Texas A&M has just established its own

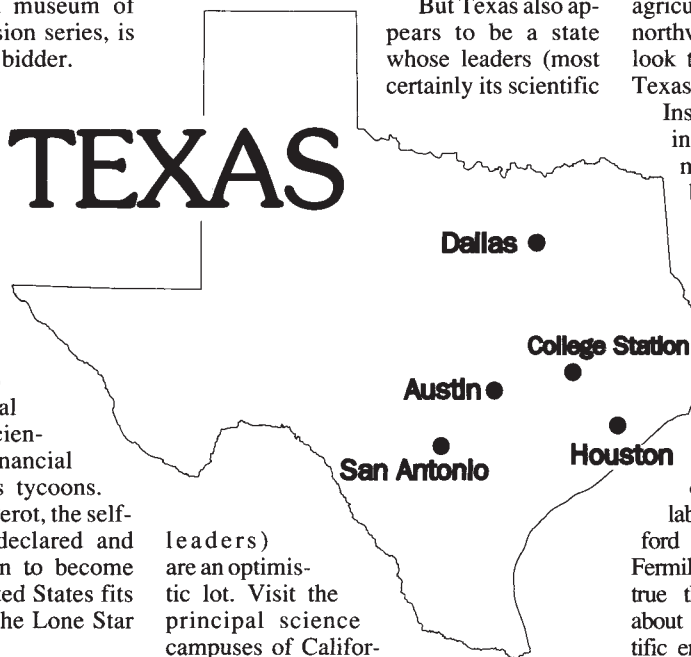
Institute of Biosciences and Technology in the midst of Houston's renowned medical institutions, where it hopes to bring its existing strength in animal biology to research that will one day apply to human medicine. In the process, the Texas A&M biotechnology centre also intends to enter the competition for grants from the US National Institutes of Health (NIH).

South of Dallas, the controversial Superconducting Super Collider (SSC) is taking shape, drawing physicists from established laboratories such as SLAC (the Stanford Linear Accelerator Center) and Fermilab outside Chicago. And while it is true that the SSC raises many questions about economic development and scientific enterprise (although it will be good for Texas's concrete businesses, it may not provide enough low-skilled jobs to have a significant impact on the regional economy), its scientific staff has adopted the optimism and boosterism that goes with Texas.

That attitude is not without some justification. Since 1981, Texas's share of the total number of US scientific papers published has risen from 5.6 to 6.3 per cent. During that period, Texas researchers have increased their output by 25.3 per cent (to 14,138 papers in 1991, according to the Philadelphia-based Institute for Scientific Information) compared to a 11.7-per-cent rise for the United States as a whole.

This growth was not, apparently, at the cost of quality. During the past decade, Texas's 'citation impact' — the average number of citations per paper — has risen 4.5 per cent, according to ISI. Even so, it is clear that Texas still has a long way to go. Despite its improvement, Texas's citation impact is still only 92 per cent of the US average.

As recently as 50 years ago, Texas was a



Texas at a Glance

Vital statistics

	US rank
Population: 16,986,000 (1990)	3rd
Size: 268,600 square miles	2nd
Gross State Product: \$340,000 million (1989)	3rd
Total R&D expenditures: \$6,576 million (1989)	6th
Academic R&D funding: \$937 million (1989)	3rd
Number of high-tech companies: 1,430 (1991)	6th

scientific backwater of the first order. Institutions that have put Houston on the medical map, such as the M.D. Anderson Cancer Center and the Texas Heart Institute, did not exist.

The Southwest Medical Center in Dallas, now home to Texas's first Nobel prize-winners (Michael S. Brown and Joseph L. Goldstein), was nothing more than a collection of army barracks from the Second World War. At the time, not a single Texas researcher was a member of the US National Academy of Sciences. Today, Texas has 42 members and, although it trails the leaders by a wide margin — California is home to 469 members and Massachusetts can boast of 255 members — Texas is proud to rank among the top ten states.

Particularly in the biomedical sciences, the early 1940s were an important time for Texas. In San Antonio, a young oil man named Tom Slick, a legend among the local citizenry, provided money to found the Southwest Foundation for Biomedical Research in 1941. The foundation has one of the most important primate research colonies in the United States and in February celebrated its fiftieth anniversary by hosting a politically contentious conference on the future of the NIH (see *Nature* 355, 573; 1992).

M.D. Anderson was also founded that year as part of the University of Texas's Houston branch. In keeping with the Texas

penchant to see the world in competitive terms, M.D. Anderson today describes its research programme, valued at more than \$90 million, as "one of the most productive targeted research efforts in the world". (M.D. Anderson ranks forty-fifth in the United States among NIH grant recipients.)

It seems that there is no scientific institution in Texas that is not building something new and M.D. Anderson is no exception. Its \$248-million building programme is grandly described as "among the largest ever in the Texas Medical Center".

Texas's competitive drive is fuelled in equal parts by homegrown pride and recognition that the state must diversify its economic base to avoid an economic collapse similar to that caused by the 1980s oil glut. Thus, winning a fierce competition to house the SSC near Dallas, and having Austin chosen as home to Sematech, the federally funded electronics consortium, gave Texas a psychological boost even if the economic payoff has yet to materialize. In the same vein, Dallas was happy to be the new headquarters of Exxon and its research and development facility. But it is in biotechnology that the greatest hope lies and where, perhaps, the most concerted effort is being made.

What follows is a look at some of Texas's important scientific institutions and their hopeful plans for the 1990s.

Barbara J. Culliton

What Texas gets from NIH and NSF

(fiscal year 1991, figures in millions)

NIH \$341.6 million 3,366 projects

Leading institutions

Baylor College of Medicine	\$73.9
UT SW Medical Center	\$54.8
UT System Cancer Center	\$35.2
UT Health Science Center, Houston	\$35.0
UT Health Science Center, San Antonio	\$34.3

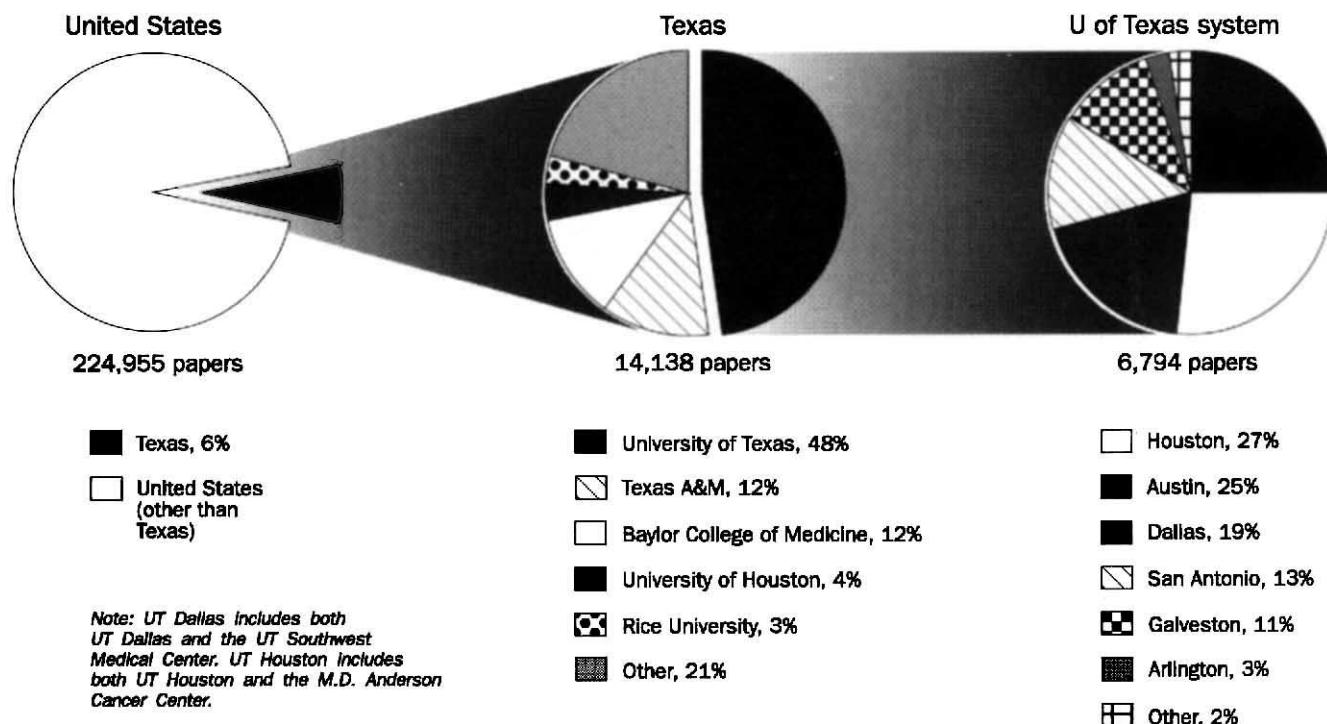
NSF \$63.4 million 711 awards

Leading institutions

UT-Austin	\$15.3
Texas A&M	\$13.8
Rice University	\$10.3
University of Houston	\$3.6

Texas science output, 1991

Total number of papers catalogued in ISI's *Science Citation Index*.



Large, oil-soaked endowment helps state universities weather current hard times

WITH the state of Texas looking at a \$5,000-million deficit this year, its higher education system is facing pressures familiar to university administrators in every other state. Texans want to support the University of Texas, says a spokesman, but invariably believe there is "fat to be trimmed". Recent talk of cutting costs by putting a limit on total enrolment, however, has alarmed taxpayers, who believe that if their money is paying for the university then their children should be able to attend it. On top of this, the university this year is facing a suit filed by the League of United Latin American Citizens and other civil rights groups who want the courts to direct a bigger fraction of education money to south Texas to redress alleged discrimination against Hispanic students.

But Texas has one thing that every other state would envy: the second largest higher education endowment in the nation, amounting to almost \$3,500 million. Only Harvard University has a bigger piggy-bank, but it does not have 15 widely scattered campuses and 150,000 students to support.

The endowment, two-thirds of which goes to the University of Texas (UT) and one-third to Texas A&M University, derives from 2 million acres of arid scrub land

in West Texas donated by the state in 1882. Ashbel Smith, the first president of UT, said at a dedication in Austin about this unpromising real estate that one should "smite the rocks with the rod of knowledge, and fountains of wealth will gush forth". In 1923, by drilling rather than smiting, oil gushed forth, and by the early 1980s UT was

that merely to maintain the buildings built during the boom may soon require capital to be drawn from the endowment.

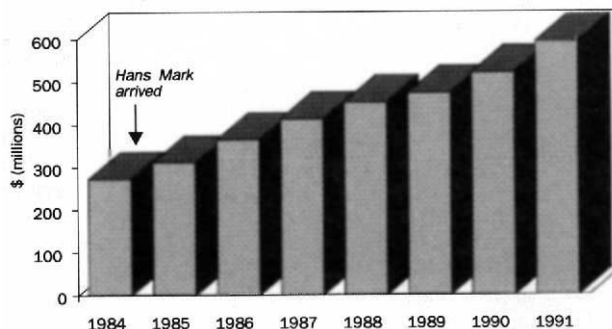
To some extent this downturn has been offset by federal research dollars, which have risen from \$273 million in 1984 to \$594 million last year. This period coincides with the tenure of Hans Mark as chancellor of the university, a former official at the US National Aeronautics and Space Administration (NASA) with plenty of federal connections. The researchers brought to Texas by oil-money in earlier years now attract money of their own from Washington, so that in the past few years overall university income has held steady. The university as a whole, says one researcher at UT-Austin, is being treated "pretty much like the average state institution", and there are plenty of places in the country much worse off.

Mark is stepping down as chancellor this year, to teach in the engineering school at Austin, and will be replaced by William Cunningham, now president of the Austin campus and dean of business administration. His academic background is presumably good experience for maintaining the health of what is by the standards of the day a thriving university system.

David Lindley

UT's rapid growth

Federal research dollars to the University of Texas since the arrival of president Hans Mark.



receiving \$200 million a year from its 'worthless' land.

This money financed a grand expansion of the university, and turned the Austin campus into a notable research institution. But lower oil prices after the oil bust of the mid-1980s and dwindling reserves mean

Academic research spending in 1991

Top Ten (excluding medicine)

Figures in \$ (millions)

Texas A&M	243
U. of Texas at Austin	208
U. of Houston	38.4
Rice University	30.1
Texas Tech U.	22.8
U. of Texas at Arlington	12.5
U. of North Texas	12.3
U. of Texas at Dallas	10.5
U. of Texas at El Paso	9.6
Southern Methodist U.	8.2

Top medical schools

Figures in \$ (millions)

Baylor College of Medicine	152
UT MD Anderson Cancer Center, Houston	108
UT Southwestern Medical Center, Dallas	94
UT Health Science Center, Houston	54
UT Health Science Center, San Antonio	45
UT Medical Branch, Galveston	38
Texas Tech U. Health Sciences Center	6.6
UT Health Center, Tyler	5.0

Source: Texas Higher Education Coordinating Board and selected institutions

Biotechnology leads the way as state moves beyond ground-based economy

Houston. Seven years after its oil industry hit rock bottom, Texas is making a comeback, and biotechnology is playing a major role in the state's effort to diversify its economy.

Leading the way is Houston — the fourth largest city in the United States — with its nucleus of about 25 companies, most of which have less than 40 employees. Central to the emergence of Houston's fledgling biotechnology industry has been the world-renowned Texas Medical Center — Houston's largest employer. The 600-acre complex, located just south of downtown Houston, is home to 41 different healthcare organizations, including the University of Texas Health Science Center, Baylor College of Medicine, the Texas Heart Institute and the M. D. Anderson Cancer Research Center.

Why has it taken entrepreneurs so long to tap into this wealth of biomedical research? Despite Houston's underlying strengths in biomedical research, it has lacked until recently the other two ingredients to any successful commercial venture — a community and culture that supports entrepreneurial activity and a local venture capital community or interest from out-of-state venture capital firms.

Although Texas has always been rich in risk capital, that money has tended to find its way back into the ground, says Martin Sutter, a general partner with the Texas-based venture capital firm, the Woodlands Venture Partners. It was generated in energy, real estate and ranching by people who understood those businesses, and reinvested the same way. Traditionally, Texans have viewed medicine and science as a philanthropic activity, says Dominic Man-Kit Lam,

professor of biotechnology, cell biology and ophthalmology at Baylor. While millions of dollars have been donated to the medical and academic communities, Lam says, "this has been done without the expectation of financial returns".

Fortunately for Lam and others, George Mitchell is a rare breed of Texas businessman. One of the things that the energy and real estate baron did, says A. Douglas Peabody, president of the New York venture firm INCO Venture Capital Management, "was to recognize that there was the potential for a third leg on the economy down there". Even before the oil business in Texas went into the doldrums, Mitchell, who is head of the Mitchell Energy and Development Corporation (one of ten major energy companies that have their headquarters in Houston), recognized an opportunity. Turning away from oil and gas, Mitchell decided to build up a high-technology and biotechnology industry in Houston.

The Woodlands

In 1974, under the management of the Woodlands Corporation, Mitchell carved out of a vacant forest 27 miles north of downtown Houston a research park and residential community. Close to Houston's Intercontinental Airport and fronting the interstate highway between Houston and Dallas, the Woodlands provides an unusual blend of businesses and residential homes within 25,000 acres of forest. Unlike some of its counterparts around the country, the Woodlands is designed to be a fully integrated community and boasts a residential population of more than 33,000, a country club, golf courses, an arts pavilion, shops, schools and churches.

One of the first facilities to be built in the

Woodlands was the Houston Advanced Research Center (HARC). Now ten years old, this nonprofit corporation links the research capabilities available through academic institutions such as Rice University, the University of Houston, Texas A&M University and the University of Texas at Austin to business, industry and the government.

Since its formation, the Woodlands has attracted a nucleus of 29 high-tech businesses employing more than 1,000 people (200–300 of whom hold PhDs). However, whereas a research park such as the Research Triangle Park in North Carolina is home to the likes of Glaxo, Inc., Burroughs Wellcome Company and Ciba-Geigy Ltd, the Woodlands Corporation has so far been unable to lure a sizeable pharmaceutical company to the park.

"We do have a hill to climb in terms of recruiting", says J. Russell Denson, president and chief executive officer of Houston Biotechnology, Inc. The problem, he says, is most acute as companies move from research into development. Lack of a home-grown pharmaceutical industry in Houston means that the city must compete with the likes of Massachusetts, New Jersey and California when trying to recruit executives with the expertise necessary to take products through to the marketplace.

Although Houston may not be everyone's cup of tea — the high temperature during the summer months can remain above 90 degrees for weeks at a time and the humidity is oppressive — industry executives hope that lack of a personal state income tax and the low cost of living in Texas will persuade researchers to move here.

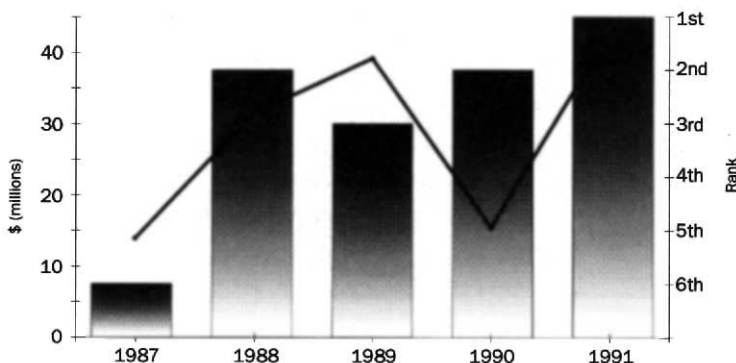
Money matters

Mitchell also recognized that Houston lacked a venture capital community of its own. In 1983, to help fill that gap, he formed the Woodlands Venture Capital Company, which became the first Texas-based fund to concentrate on biomedical ventures. Before that, says Sutter, "there really wasn't an institutional venture capital firm in Texas that did early-stage medical ventures".

At about the same time that Mitchell established the Woodlands Venture Capital Company, BCM Technologies, Inc. was formed as the technology transfer arm of Baylor College of Medicine, a private school in Houston. A biomedical start-up called Cardiovascular Systems, with backing from the Woodlands Venture Capital Company, was one of the first companies to be launched from an academic institution in the area. Founded in 1985, the company's core technology was an autologous blood transfer system developed under licence from

Venture capital financings

Line represents venture capital financings for medical/health-care sector (1987–1991) in which Texas venture capital firms participated, and bars show the rank of the medical/health-care sector in relation to other sectors of industry.



Source: The Texas Venture Report 1991 Annual Survey, Deloitte & Touche, Dallas, Texas.

Baylor. The company shipped its first product within a year, became profitable one year later and was sold within three years, netting a tidy profit for the company's founders and for Baylor. This early success, Sutter says, caught the attention of other Texas-based venture groups, including the Perot Group, Interwest and Triad Ventures Limited, all of which now have an active interest in biomedical ventures.

Although the Woodlands Venture Capital Company and the newer independent venture fund, the Woodlands Venture Partners, have been instrumental in taking the lead in new venture deals, much of the supplemental capital comes from out-of-state investors. Since 1983, the two venture funds have invested \$22–30 million in 18 companies, including Houston Biotechnology, Inc., LifeCell Corporation, Argus Pharmaceuticals, Inc. and Triplex Pharmaceuticals Corporation. All but two of the 18 companies are related to health care, Sutter says.

The statistics also support this growing interest in the medical/healthcare sector. In 1991, for the first time in five years, the medical/healthcare industry sector garnered the largest share of Texas venture capital — 29 per cent, or \$43 million out of a total of \$148 million (see chart).

Local and state officials are also jumping on the biotechnology bandwagon; both the governor, Ann Richards, and new mayor, Bob Lanier, have taken up the cause. Assisting with these efforts is the Texas Department of Commerce in Austin, under the direction of Cathy Bonner. While department officials have recognized for some time that Texas must depend less on its natural resources, it is only in the past year that the department has singled out biotechnology as one of six emerging technologies of economic importance to the state of Texas.

To understand why this is so, Michael Klonsinski, programme director at the Commerce Department's Office of Advanced Technology, says it is important to remember that the department is only five years old. During its boom years, Texans never felt a need to promote economic development.

Klonsinski's office plans to provide financial support on a modest scale to companies in the targeted industries. Starting this autumn, his office will make up to \$25 million available to companies to fund the development of new products.

In the past year, several of Houston's biotechnology companies have made public stock offerings. At least three more are watching the stock market in the hope of going public sometime this year. But what will really place Houston on the biotechnology map, Peabody says, is for one of its start-ups to market a drug that starts generating revenue and profits that can then be reinvested in the business. With a handful of companies with drugs in clinical trials, that may happen sooner rather than later.

Diane Gershon

SW medical center, famous and favoured, is thriving

Dallas. When Michael S. Brown and Joseph L. Goldstein won the Nobel Prize in 1985, all of Dallas celebrated. Texas takes competition seriously. Goldstein and Brown, who brought Texas its first (and so far only) Nobel Prize, for their research at the University of Texas Southwestern Medical Center on cholesterol metabolism and its genetic control, are local heroes still. And the medical school to which they have devoted their careers and that was once a scientific no man's land is thriving.

At a time when many US medical schools are feeling gloomy, pinched for cash and blocked from expanding, Southwestern is optimistic and remarkably well-to-do. It is a favourite charity of Dallas's millionaires and billionaires, including US presidential hopeful H. Ross Perot, who is often seen at the hospital cafeteria recruiting leading researchers to Texas.

Perot recently donated \$15 million for an MD–PhD programme that Brown and Goldstein have been eager to start. And this past December, the university announced four private gifts totalling \$85 million. Dallas philanthropist Nancy Harmon gave \$25 million, as did an anonymous donor. The Excellence in Education Foundation (established by the co-founders of Texas Instru-

ments, Inc.) sent a cheque for \$30 million, and the Southwestern Medical Foundation contributed \$5 million. Much of the money will go to Brown and Goldstein's research on molecular genetics, as will some of an additional \$100 million that has been pledged by the Texas state legislature.

Southwestern is also rapidly expanding its physical plant and creating a new campus. Six new research buildings (one already going up and a second on the drawing board) are planned for a 39-acre site just north of the medical centre campus, which will essentially double Southwestern's physical size. An elevated tramway will connect the two campuses.

In many cities, such expansion would be a medical dean's nightmare. Neighbours now routinely challenge an academic institution's plans to expand, and environmentalists might be expected to object to building an elevated highway. But not in Dallas.

According to Kern Wildenthal, president of Southwestern, the local community is proudly supportive of its medical centre, with city officials and businesses eager to help make Southwestern the best there is. Or, as university vice-president William Neaves puts it, "there is a healthy underdog atmosphere here that makes the school and the city pull together."

The John D. and Catherine T. MacArthur Foundation of Chicago gave Southwestern's dreams a real boost when it donated 30 acres as part of its own 'asset management' strategy. The foundation owns another 90 acres just across the intersection from the medical centre, and it hopes that Southwestern's new research buildings will form the core of a new research park on land that MacArthur will sell at market value — when the market is high enough.

Southwestern's current stature and gleaming new laboratories are a far cry from the institution that came into being 50 years ago next year. Baylor College of Medicine, originally a



Michael Brown, left, and Joseph Goldstein provide a foundation for the center's growth.

Dallas institution, moved to Houston in the early 1940s, where it has grown into an important clinical research centre. And although it did not look auspicious at the time, it turns out that Houston's gain was Dallas's gain as well.

In 1943 the Southwestern Medical Foundation founded the medical school to take Baylor's place. For the first few years, a homegrown student body learned medicine from a small faculty and practising physicians in Dallas who spent part of their time teaching. Then, in 1950, Southwestern recruited the man who, during the next four decades, would turn a ramshackle school into a major medical research center.

Donald Seldin was on the faculty of medicine at Yale ("a crowded department

that didn't offer me much opportunity to do anything fresh," he recalls) when he was recruited to Dallas. "I was so naive," says Seldin, a native of Coney Island, New York, "that I didn't even visit. We just arrived, sight unseen, in January 1951. It was quite a shock."

The medical school was housed in old army barracks that had been a staging area for soldiers moving on to San Antonio, "wooden shacks, really, with no foundation, just sitting there on top of the ground", Seldin recalls. "The pipes were lying on the ground and when it got cold and the water froze we had to close the place down." There were no research facilities of consequence and Parkland Hospital (subsequently made famous when John F. Kennedy was taken there after being struck by an assassin's

bullets) was "out of date before it was even built".

Seldin's concern about having to replace existing faculty was unfounded, for when he arrived he discovered that "there were no faculty to displace". For someone who wanted a chance to do something fresh, Seldin was in the right place. Starting with the philosophy that "biomedical science is the foundation of medicine", he organized the first fellowship programme at Southwestern and encouraged Southwestern graduates to go out into the world for training. Goldstein, for instance, trained at the Massachusetts General Hospital and the National Institutes of Health before heeding Seldin's call to return to establish a department of genetics. His partner would be Brown, whom Seldin had lured from the University of Pennsylvania.

The intellectual history of Southwestern has been one of gradual development that continues to this day. In the 1960s, Seldin put the foundation of a science-based medical curriculum in place to support clinical teaching. Cell biology, immunology and transplantation immunology were among the departments that took root. In the 1970s, Brown and Goldstein's programme in molecular genetics flourished, along with the growth of the department of biochemistry. Molecular biology had its turn in the early 1980s, followed by attention to building a strong base in the neurosciences with the recruitment of James Hudspeth from the University of California at San Francisco.

In keeping with its tradition of making steady progress, Southwestern in the 1990s aims to strengthen its clinical departments (perhaps in competition with its medical rivals in Houston). Southwestern officials believe they have added a feather to their cap by attracting John Minna from the National Cancer Institute to develop a strong programme in clinical oncology and basic cancer biology.

Altogether, Southwestern seems still to be a very up-and-coming place, blessed with land, Texas money and, its department chairmen say, an unusual atmosphere of goodwill. "We are all actually friends," says one, to which a colleague adds, "It is easy to be friendly and cooperative when you have lots of resources as we do."

At present the Texas state legislature is reassessing the measures it uses to allocate funds to state institutions, a prospect that Southwestern seems not to fear at all. Among the possible yardsticks are the number of medical graduates per school, the number licensed and the number certified by medical specialty boards. The percentage of charity care in the hospital is also on a potential list of measures, as is a quantitative assessment of the amount of research a school does. Says Southwestern's Wildenthal, "by whatever measure, Dallas will do well."

Barbara J. Culliton



Southwestern Medical College in the 1940s.



The University of Texas Health Science Center at Dallas today.

Houston breeds biotechnology hot shots

Houston. In 1984, Dominic Man-Kit Lam and Jared M. Emery discovered a monoclonal-antibody-based product to prevent the recurrence of cataracts after surgery. Lam says that he and Emery, both professors of ophthalmology at Baylor University, had intended to license the product to an out-of-state pharmaceutical company. But after discussions with James Elkins Jr, then chairman of First City Bank and vice chairman of the board of trustees at Baylor, and energy and land baron George Mitchell, they were persuaded to keep their invention in Texas.

With that, Lam and Emery founded Houston Biotechnology, Inc. (HBI), one of the few companies to concentrate on developing novel therapeutic agents with ophthalmic applications. In 1986, HBI became one of the first biotechnology companies to establish its headquarters in the Woodlands — the 25,000-acre high-tech community outside Houston.



Dominic Man-Kit Lam.

HBI's flagship product, known as 4197X-RA, is an immunotoxin (monoclonal antibody conjugated to a plant toxin) designed to destroy selectively any lens epithelial cells remaining after cataract surgery. Epithelial cells that are not destroyed cause cataracts to re-form in about half of the patients that have undergone surgery. J. Russell Denson, HBI's president and chief executive officer (CEO), hopes to have the drug on the market by 1997 and estimates that annual worldwide sales could exceed \$600 million. HBI has already caught the eye of the leading Japanese ophthalmic company, Santen Pharmaceutical Company Ltd, which holds exclusive rights to market HBI's drug in Japan.

Since founding HBI, Lam has established two other companies — LifeTech Industries Ltd and GES Pharmaceuticals. Last year, he was appointed chairman and CEO of AgriStar, Inc., one of only a handful of start-up agricultural biotechnology companies in Texas. In the year since his appointment, Lam has brought the company back from the brink of bankruptcy and watched its market valuation rise from a low of \$2 million to almost \$40 million. At the heart of the company's technology is a patented membrane system for the cloning and micropropagation of plants. The membrane lets gases and light through but is impermeable to liquids and bacterial contaminants.

So far, Lam has resisted the temptation to license the technology to the giant agricultural companies. He says AgriStar prefers instead to undertake custom projects, and last March it agreed to micropropagate a proprietary line of potatoes for the Monsanto Company.

Triplex Pharmaceutical Company was formed three years ago as a spinoff of Michael Hogan's laboratory at Baylor. This Woodlands-based company is developing human therapeutics based on triple-helix-forming oligonucleotide compounds that bind directly to DNA and inhibit expression of protein at the level of transcription. Triplex hopes to be able to apply the 'triple-helix' approach to any disease where there is abnormal or unwanted gene function, such as is seen with certain cancers, cardiovascular and autoimmune diseases. But the most immediate application of the technology is in the treatment of human immunodeficiency virus (HIV) and herpes simplex virus. Candidate compounds are being evaluated in animal studies.

With \$19 million in venture financing under its belt, Triplex earlier this year signed a multi-year collaboration with Hoechst AG, one of the world's largest pharmaceutical companies. The deal, worth \$30 million, grants Hoechst worldwide marketing and manufacturing rights to antiviral drugs developed by Triplex using the triple-helix technology.

LifeCell Corporation was formed in 1986 to commercialize an ultra-low temperature freeze drying technique for biological cells and tissues, a technology developed at and licensed by the University of Texas Health Science Center (UTHSC) in Houston. It was the first time that any part of the UTHSC system retained an equity position in a company. LifeCell spun off its equipment business to RMC, Inc. in 1991 to concentrate on the therapeutic applications of its core technology.

LifeCell has six government grants and contracts worth almost \$2 million, with funding provided by the US Navy, the US Army and the National Institutes of Health. In addition, LifeCell recently raised \$7.1 million in an initial public offering. Clinical trials are planned for July with the company's leading product, a transplantable full-thickness skin graft using processed human cadaver skin. The skin graft is designed as a permanent treatment for patients with third-degree burns and bed sores.

LifeCell's tissue-processing technology selectively removes the cellular component of skin that elicits an immune response and causes graft rejection, leaving the structural dermal matrix and basement membrane intact to serve as a template for repopulation with the patient's own cells. The company hopes to have the product on the market next



LifeCell Corporation's facility in the Woodlands research park, Houston.

year. Additional transplantable tissue grafts under development include vascular conduits for bypass surgery and replacement heart valves.

In a southwest section of Houston, the wife-and-husband team of Nancy and Tse Was Chang have been quietly building a biotechnology company. Founded in 1986 on the basis of technology transferred from Baylor, Tanox Biosystems, Inc. is developing neutralizing monoclonal antibodies for the treatment of HIV, allergies and certain autoimmune diseases. And, while the company maintains a low profile, it has caught the eye of the Swiss pharmaceutical company, Ciba-Geigy Ltd. Collaborative research agreements between Ciba-Geigy and Tanox cover the development of the company's lead product, CGP47439, for the treatment of AIDS (which is in early-stage clinical trials) and the company's allergy research.

Nancy Chang is no stranger to the biotechnology industry, having worked for the now ailing Centocor, Inc. She believes that the right way to build a company is slow and steady. "It is not what you say, it is what you hope you can do", she says. "Eventually you've got to deliver."

One of Houston's most recent biotechnology start-ups is Texas Biotechnology Corporation. Located in the heart of the Texas Medical Center campus, the company is developing novel biotherapeutic agents to treat cardiovascular diseases such as atherosclerosis and hypertension. However, unlike most of the Houston-based start-ups, it is not a product of just one academic institution. Texas Biotechnology has close associations with several institutions, including the UTHSC, Baylor, Texas Heart Institute and the M. D. Anderson Cancer Research Center. Richard Dixon, scientific director at Texas Biotechnology and former head of molecular biology at Merck Sharp and Dohme, says that last October the company raised \$22 million from private donors to finance its research.

Diane Gershon

New biotechnology institute at Rice promotes interdisciplinary studies

LAST November, Rice University opened a state-of-the-art biotechnology research and training facility that brings together three of the university's laboratories — the Cox Laboratory for biomedical engineering, the Greenwood Laboratory for basic medical science and the Mabey Laboratory for biochemical and genetic engineering.



Institute of Biosciences and Bioengineering — Rice's new interdisciplinary research and educational facility.

The Institute of Biosciences and Bioengineering is a three-storey, 108,000-square-foot facility housed in the George R. Brown Hall. Its mission, explains its assistant director, Diana Welch, is to promote interdisciplinary research activities between faculty members, students and staff at Rice and the neighbouring Texas Medical Center.

In addition to research, Welch says, the institute hopes to train the next cadre of scientists to feed into the state's burgeoning biotechnology industry. The institute offers a graduate training programme in biotechnology, which includes placing students at a biotechnology company for 3–6 months.

The \$25-million facility was funded largely through philanthropic support. Of the \$15 million in private donations, \$2.5 million came from the George R. Brown Foundation. Brown is a Rice alumnus and former chairman of the university's board of governors.

In addition to the 27 Rice faculty members, the institute has 12 adjunct faculty from Baylor College of Medicine and the University of Texas Health Science Center at Houston. Plans are afoot to recruit seven additional faculty members in biochemistry

and cell biology and two in chemical engineering.

In January, Antonios Mikos thought enough of the institute's potential to leave the Massachusetts Institute of Technology (MIT) and become an assistant professor of chemical engineering and bioengineering in the Cox Laboratory. Mikos was lured by a research environment where biologists, chemists and engineers can freely interact, and close proximity to the Texas Medical Center.

In collaboration with researchers at MIT and Harvard, Mikos is developing a method of regenerating fully functional liver tissue by growing liver cells on a bioresorbable, three-dimensional polymer scaffold. If successful, the procedure would provide an alternative to whole organ liver transplants and would alleviate a critical shortage of donor organs. Once implanted, the scaffold becomes innervated with blood vessels and provides a support matrix that can then be seeded with liver cells that have been cultured *in vitro*. The new liver transplanting technique has shown promise in small animal models and is being tested in pigs. Mikos hopes to begin human clinical trials within two years.

Diane Gershon

Texas's top twenty fields

Science subfields in which Texas papers get more citations than the US average (= 1.00) in that field.

1	Physics	1.23
2	General Clinical Medicine	1.19
3	Orthopaedics/Traumatology	1.16
4	Medical Technology	1.16
5	Gastroenterology	1.15
6	Otolaryngology/Ophthalmology	1.14
7	Agriculture/Agronomy	1.13
8	Oncology	1.13
9	Life-Sciences Chemistry	1.12
10	Paediatrics	1.12
11	Instrumentation	1.09
12	Environmental/Social Medicine	1.09
13	Haematology	1.08
14	Biotechnology/Applied Microbiology	1.07
15	Analytic, Inorganic & Nuclear Chemistry	1.06
16	Organic Chemistry	1.05
17	Cardiology & Respiratory Medicine	1.05
18	Experimental Biology/Medicine	1.03
19	Chemistry, General	1.03
20	Neurology	1.02

Source: ISI's Science Indicators Database, 1987-91

The woods are alive with high technology



On 100 acres in the 'research forest' of the Woodlands, the Houston Advanced Research Center (HARC) has become a focus of Texas physical sciences and technology research. Started by a Texas oil millionaire ten years ago, it has grown from a consortium of four Texas universities to a joint centre involving eight US universities and the Hungarian Academy of Sciences. Its 140 researchers and technicians do materials research, astrophysics, laser research and are developing magnets and beam technology for the Superconducting Supercollider as part of a \$14 million research programme.

San Antonio fills research park by making offers that scientists cannot refuse

San Antonio. In the middle of nowhere on the outskirts of San Antonio, a new research park is in the making. The Institute of Biotechnology is first on the prairie.



Ross Perot, left, with Jess Hay, former chairman of UT board of regents, and John Howe III, president of UT-Health Science Center in San Antonio. At right, Wen-Hwa Lee.

was so committed to biotechnology that it built the laboratory before anyone had agreed to fill it. Acting in the belief that the city could prosper economically by capitalizing on its many medical research institutions (see story on page 638), the city had plunged boldly into a hoped-for but nonetheless uncertain future. An oil man named Tom Pawell donated 1,500 acres for a research park and the university pledged \$30 million for equipment and research to the new Texas Research and Technology Foundation. Perot wrote a cheque for \$15 million to pay for construction of the limestone and glass laboratory named after a local entrepreneur, Hayden Head.



Lee was also promised close ties with the medical centre's department of medicine, where chairman Jay Stein was eager to forge a collaboration with Lee's research group, which focuses on cancer genes and regulatory problems.

Lee was interested but not convinced. There was, he told Howe, one thing missing. Houses. Out on the prairie there was no place to

live. "We work 24 hours a day," Lee says. "This is a very competitive field. Post-docs and graduate students come and go at all hours. They cannot have a long commute. It is not efficient."

No problem. A two-storey complex with 30 apartments, enclosing a garden with swings for the children, has been built within walking distance of the laboratory, and more faculty housing is on the drawing boards.

In July 1991, Lee and his colleagues from San Diego moved into Hayden Head. The faculty now is up to ten, with researchers from Princeton, Johns Hopkins and Dartmouth universities among those who have been enticed by the newness of it all and the charm of watching 60 deer eat dinner at the institute's feeder.

Scientists who are there are excited by the opportunity to build a research laboratory from scratch. However, they miss the well-stocked libraries that are the hallmark of a major research institution and they feel that their relations with medical colleagues are constrained by the fact that the hospital is miles away in downtown San Antonio.

Will this new venture work? Although it is anybody's guess, those involved, whether Texas-born or newly adopted, have no doubts.

Barbara J. Culliton

"Young man," said Ross Perot to Wen-Hwa Lee, "I want you to come to Texas. You tell me what you want. Is that fair enough?" Lee, then head of a molecular biology laboratory at the University of California at San Diego, thought it was. Says Lee, "Mr Perot said that if I came to Texas and needed anything, I should just let him know. I found that very encouraging."

Perot, a Texas legend long before he launched a campaign for the US presidency, has played an active part in recruiting scientific talent to Texas, particularly in San Antonio and Dallas. But Perot's open-ended offer to Lee was only one of many elements that induced him to move his 23-person laboratory from San Diego to San Antonio.

Lee and his wife, Eva, grew up in Taiwan. Each migrated to California for a PhD at the University of California at Berkeley before moving to San Diego. Although his laboratory was successful, Lee saw more clearly the limits to where he was once the president of the University of Texas Health Science Center, John Howe III, began to recruit him in 1990 to San Antonio.

For instance, he says, "We had no room. We had 23 people in 2,000 square feet." Howe was offering him a brand-new research building just waiting to be occupied.

Lee was also impressed that San Antonio

Texas' top ten turnarounds

Fields whose citation "impact" have gone from below to above the US average (= 1.00) in the last decade

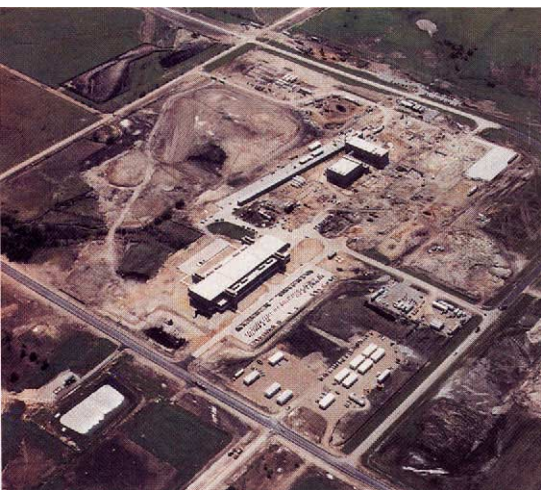
Citations per Texas paper relative to US average

		1981-85	1987-91	Change
1	Biotechnology/Applied Microbiology	.63	1.07	+.44
2	Otolaryngology/Ophthalmology	.71	1.14	+.43
3	Physics	.86	1.23	+.37
4	Gastroenterology	.82	1.15	+.33
5	Environmental/Social Medicine	.81	1.09	+.28
6	Orthopaedics & Traumatology	.90	1.16	+.26
7	Cardiology & Respiratory Medicine	.83	1.05	+.22
8	Organic Chemistry	.88	1.05	+.17
9	Haematology	.92	1.08	+.16
10	Neurology	.91	1.02	+.11

Source: ISI's Science Indicators Database, 1981-91

Supercollider creates a field of dreams for luring high-energy physicists to Texas

Dallas. Texans like to think that their state contains the biggest and the best of everything. And there is no denying that, even by Texas standards, the Superconducting Supercollider (SSC) is big. But six years after former president Ronald Reagan announced his support for its construction and despite a proposed budget of \$650 million for 1993 and nearly 2,000 people working in temporary offices at two sites 15 and 25 miles south of Dallas, serious questions remain about the SSC's chances of becoming the most powerful proton-antiproton collider in the world.



The magnet testing site, including a 265-foot-deep shaft (top) that will allow tunnel boring machine to begin its work.

Can the United States really afford it? Can it be built on schedule and within budget? And, even assuming that it meets its timetable and is turned on in 1999, will the 40-TeV collider do the physics that scientists expect?

Fred Gilman is confident that the answer to each question is yes. After working for 23 years as a theoretical physicist at the Stanford Linear Accelerator Center (SLAC) in Palo Alto, California, Gilman joined the SSC laboratory two years ago because he was convinced that Texas had become "the most exciting place for a high-energy physicist to be". Now, as head of physics research, Gilman spends a good deal of his time making that point to his colleagues around the world.

A few years ago, a popular film called "Field of Dreams" described how an Iowa farmer turned his fantasies into reality by building a baseball diamond in his cornfield and attracting star players from the past. Its

most memorable line was, "If you build it, he will come".

For Gilman, however, the opposite is true. He has bet his career on the notion that some 165 additional PhD physicists over the next several years, an average of two a month, will come to the suburbs of Dallas to help him build and operate the world's most powerful particle accelerator. If he is right, their presence and \$8,250 million from the US government, the state of Texas and as many foreign contributors as possible will help to turn Gilman's own 'field of dreams' into scientific reality.

"I just couldn't pass up the opportunity to build something from the ground up", says Gilman, a member of the original advisory group that reviewed designs for the accelerator, in explaining why he left SLAC for Texas. "Something spectacular is going on here, but sometimes, when I step back, it's hard to believe that it's really happening."

Gilman's enthusiasm is typical of the attitude among SSC scientists. A common thread drawn from numerous interviews is their pride in being part of something larger than themselves, of tackling problems that dwarf previous challenges. To a man — there are enough women at the SSC laboratory to form a 30-person chapter of Women In Science and Engineering (WISE) but too few to make such a group unnecessary — the scientists hold to the view that the United States can afford the SSC, that the money will be forthcoming and that the project will work.

They are sceptical, if not dismissive, of efforts by Carlo Rubbia, director general of CERN (the European Laboratory for Particle Physics), to build a Large Hadron Collider by the end of the decade under the Jura Mountains on the Swiss-French border. They believe that CERN will have problems finding the money for its next big machine, and that its scientists must overcome daunting technical problems in building the superconducting magnets to specification. Voting with their feet, they remain convinced that Texas is where the next great secrets of the fundamental components of nature will be revealed.

For some, the thrill is in the construction phase. "I like to build big toys for scientists", says Ted Kozman, an engineer who heads the accelerator systems divisions and has worked on major construction projects at Lawrence Livermore and Lawrence Berkeley national laboratories, "and I couldn't imagine too many more opportunities in my lifetime."

"When Dick Briggs [deputy director of the laboratory and a former associate director at Livermore] asked me if I was inter-

ested in coming to the SSC, he didn't have to twist my arm."

How Big?

The scale of the SSC is unlike any previous scientific project. In addition to pumping millions of dollars into the Texas economy, consider that:

- There are actually five distinct accelerators being designed and built at the laboratory that will, step by step, boost the energy levels of the circulating protons and antiprotons to the point where they will collide with unprecedented frequency. One of those injectors, called the medium-energy booster, will upon completion in 1996 be the third most powerful accelerator in the world.

- More than 1,500 scientists at universities, national laboratories and companies around the world will be involved in designing and building the two initial detectors to be installed at the SSC. Each detector will cost more than \$500 million, weigh about 50,000 tonnes, contain millions of miles of electronic wiring and be capable of recording up to 300 million collisions per second. The detector collaboration known as GEM (Gamma, Electron and Muon), for example, will feature the world's largest superconducting magnet, a horseshoe-shaped object 17 m in width.

- The first test of a half-string of superconducting magnets (five 16-meter-long dipole magnets and one quadrupole magnet) for the main accelerator will be conducted in the autumn. If the magnets perform as expected, two industrial consortia will begin mass production of a total of 8,600 dipole and 2,000 quadrupole magnets to be distributed evenly around the 54-mile elliptical tunnel that will house the SSC. By comparison, the Tevatron at Fermilab in Illinois, currently the world's most powerful proton accelerator, contains a total of 1,000 superconducting magnets.

Yet for all its impressive numbers, the most striking feature of the SSC laboratory is the fact that an enormous work force — comprised of scientists and secretaries, technicians and accountants, managers and maintenance crews — is working side by side to create something out of nothing. It is a sobering reminder that big science still requires the successful efforts of thousands of little people.

"Sure we wish we had more money", says Roy Schwitters, a former Harvard University physicist who headed the team for the Central Detector Facility that was installed in the mid-1980s at Fermilab and who, in 1989, was appointed as director of the SSC laboratory. "But brains, commit-

ted to the long haul, are our most vital resource."

The unprecedented size of the laboratory affects every aspect of the project. A steady stream of organizational changes reflect the fact that, as Schwitters readily admits, "it's a harder job that I expected. I had no idea of the demands from the business side of the equation."

There have been more than half-a-dozen project managers since the laboratory became the first tenant in a business park owned by Texas real estate magnate Trammell Crow. Schwitters more than once reshuffled his own deck of top administrators, and in October 1990 his bosses at the US Department of Energy decided to appoint Edward Siskin, a nonscientist with experience in running large construction projects, to the new position of general manager. At the same time, Schwitters inducted Paul Reardon, a veteran accelerator builder, to join the laboratory as project director.

While Siskin remains in charge, the supporting staff continues to evolve. Last July, when accelerator division chief Helen



The linear accelerator is the first of four machines that will prepare protons for the SSC.

Edwards left, her job was divided between Dugan and John Rees, who came from SLAC. In December, her husband and associate, Don Edwards, also retired from the laboratory. In January, Reardon went on medical leave.

In the wake of those departures, the entire accelerator division became part of project management. Rees succeeded Reardon as project manager, and Dugan added the job of deputy project manager to his duties as head of the accelerator division. Such turnover may not be unusual for an organization as large as the SSC project, but it has drawn increased attention to a project already under intense scrutiny because of its cost.

If the SSC succeeds, Texas will become the centre of the universe for high-energy physics. Its failure could be equally momentous — the end, perhaps, of an era in which the United States has led the world in big science.

"We're breaking new ground here", says Schwitters, "And even the old hands in the field — people like Robert Wilson [Fermilab's first director] — admit that the SSC is a different world." **Jeffrey Mervis**

Four who have come

As with any new laboratory, the people working at the Superconducting Super Collider (SSC) have come from somewhere else. Their presence, in many ways, is a testament to their faith that the huge collider will not only be built, but will also be worth working on.

Fred Gilman, head of physics research, and Jerry Dugan, who oversees the science and the construction of the big accelerators, are part of a large group who have left existing high-energy physics laboratories in the United States. Gilman had spent his entire career at the Stanford Linear Accelerator Center, and Dugan worked at Fermilab for eight years, before both joined the SSC in 1990.

In addition, a significant number of researchers — some 30 per cent of his scientific staff, Gilman estimates — are from laboratories outside the United States, in particular DESY in Germany and CERN, although some of those are Americans returning home. Still others are on short-term contracts, notably a contingent of Russian high-energy physicists from the scientific complex at Novosibirsk. And there are a growing number of professors from US universities, both nearby and around the country, who will spend varying lengths of time at the SSC laboratory working on specific projects.

They come to Texas for all the usual reasons: the opportunity to make their mark, to assume greater responsibility, to tackle the next great challenge. For the average Texan, the SSC represents a \$2,000-million construction project, its 54-mile elliptical tunnel providing thousands of jobs for an economy just beginning to climb out of a long and severe recession. But for high-energy physicists, the SSC is the future.

Here is a brief look at four who have taken the plunge:

■ After spending almost 20 years at Fermilab, Sam Baker jumped at the chance to join the SSC two years ago because "it is such a wonderful opportunity to start from scratch". A health radiation physicist who directs the growing environmental safety team at the laboratory, Baker also finds time to carry out his own research projects as well as to make regular visits to area schools.

"Educating the next generation is part of everybody's job here", he says about his community involvement. "I can't imagine not doing it, it's so important." At the age of 57, Baker expects to retire from the laboratory after the collider is turned on and the environmental programme is fully in place.

■ Weiren Chou became an accelerator physicist by accident in 1986, when he joined Argonne National Laboratory outside Chicago to study nonlinear dynamics of the giant machines. Chou came to the United States in 1980 as part of the pioneering efforts of Nobel laureate T. D. Lee of Columbia University to attract talented graduate students from mainland China.

Although Chou says he was very happy working on the Advanced Photon Source being built at Argonne, he says SSC officials contacted him at a national physics conference and persuaded him to come to Texas in 1990. His only complaint is the lack of established procedures for purchasing things, large and small: his group has tried for months to buy a \$50,000 computer work station, and he has almost abandoned hope of getting the laboratory to pay for business cards to give to vendors.

■ Vera Luth's arrival this spring from SLAC is a measure of how far the laboratory has come. Two years ago Luth, an experimental physicist, decided she was not ready to be one-quarter of the scientists in the physics research division. But the opportunity to be Gilman's deputy was sufficiently attractive to lure her and her husband away from what she calls "probably the best place to live in the United States".

Luth has hedged her bet; her husband and she decided to rent rather than buy a home, and she is officially on leave from SLAC to give her time to settle into her new job. So far she is withholding judgement on the laboratory. She is concerned about who will pay for the two huge detectors and whether they have grown too big for university teams to handle, and she is unhappy that such megaprojects take so long to design and build that they have knocked students out of the picture. But she says that those are facts of life for today's high-energy physicists, and she seems eager to find answers to these and other problems.

■ A restructuring of the SSC laboratory last year gave Jerry Dugan the chance to walk away from operating a 'mature' accelerator at Fermilab and embrace the challenge of building new and bigger accelerators. It was the second time in two years that he had seized an opportunity created by his boss at Fermilab, Helen Edwards. Edwards left Fermilab in 1989 to join the SSC project; last July, she retired and returned to Illinois.

Such musical chairs is common in the field. Dugan is hoping to fill the remaining 20 or so slots in his programme with experienced accelerator physicists, which means persuading them to forsake comfortable positions at existing laboratories.

J.M.

Texas A&M struggles to gain the respect that its researchers have already won

College Station. On a vast plain 100 miles northwest of Houston, Texas A&M University remains one of the best-kept secrets in science. Although it consistently ranks among the top ten US research universities by such measures as total research spending and science degrees granted, to many it is best known as an agricultural school with a good football team.

Such is the lot of a land-grant university. Texas A&M, as dozens of other universities in agricultural states, was founded on land donated by the US government for the purpose of agricultural education. Over the decades, such universities became arms of the US Department of Agriculture (USDA), receiving virtually guaranteed federal funding and focusing on their service-orientated agricultural mission.

Without abandoning its commitment to the Texas farmer, Texas A&M has grown up in a hurry. In 1969, it admitted women for the first time and dropped its military-school trappings. Between 1970 and 1980, the university grew nearly fourfold; it now has more than 41,000 students, the seventh highest enrolment in the United States.

With the increased student population came new money, and expansion into research disciplines as diverse as oceanogra-

phy and space. Texas A&M operates one of the largest university fleets in the nation, including an ocean-drilling ship and a converted ocean liner that is used as a classroom. And the National Aeronautics and Space Administration (NASA) designated it as one of its 'space grant' universities, which provides funding for a permanent space studies department.

But Texas A&M's growth has now slowed. The university has essentially capped enrolment, which means that state funding has also reached a plateau. "We don't have an expanding student population any more, so asking the state for more money is difficult," says Charles Arntzen, a Texas A&M biochemist. That goes for USDA funding as well. The department is shifting its research funding into competitive grants, not formula-based entitlements to land-grant universities. If Texas A&M wants to stay near the top of the list of federally funded universities, it must now compete for grants like the rest.

Fortunately, Texas A&M has an ace up its sleeve — its endowment. Some apparently worthless land given to Texas A&M and the University of Texas long ago by the state turned out to hide a vast oil reservoir; its one-third share makes Texas A&M the

seventh richest university in the country. That kind of financial security has allowed it to invest in the kind of research infrastructure that can attract outside funding.

Combining university funds with private money has enabled research at Texas A&M to grow even when traditional funding sources have not. A \$27-million Institute of Bioscience and Technology (IBT) in Houston is almost ready (see story on next page), and work is about to start on a complementary Crop Biotechnology Building on its main campus here in College Station as part of a planned \$40-million Center for Southern Crop Improvement. In general, the endowment helps to pay for new buildings, while grants and private funding help to fill them.

But as a state university, Texas A&M must answer to the legislature as well as to its own faculty. Some lawmakers do not agree with the university's emphasis on research, which they fear is done at the expense of education. Many Texans, whose children are rejected because of new limits on enrolment, think that the university's endowment might better be spent on accommodating more students.

"We're concerned that if we're too elitist, the legislature will change the constitu-

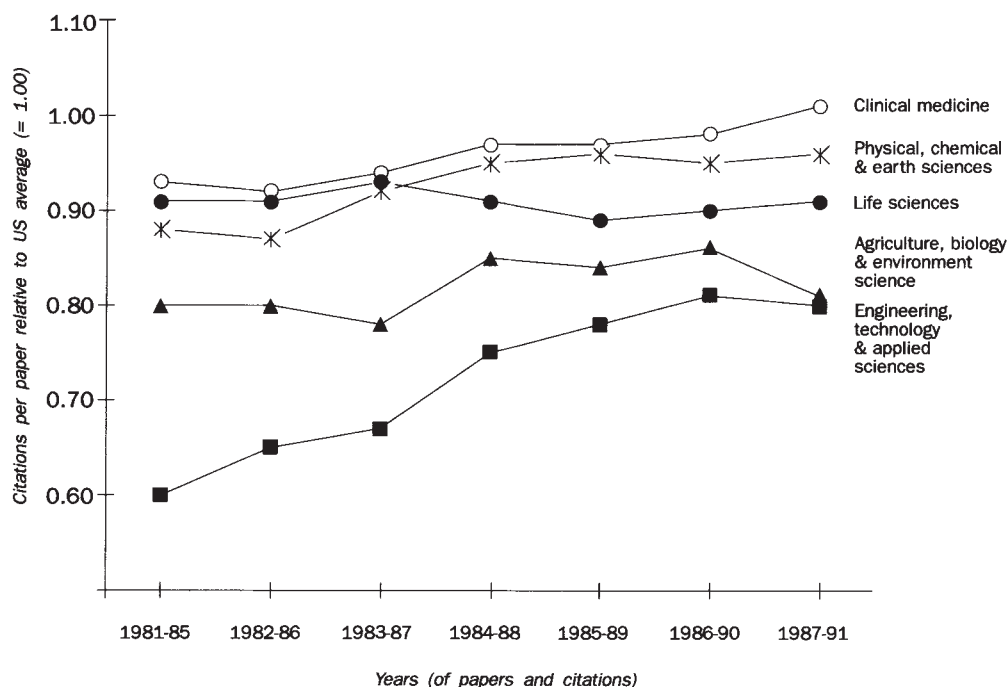
tion to make the money available to other universities," says John Shaddock, dean of the college of veterinary medicine. One solution is to explain to the public the value of research. The public, in this case, includes the governor, Ann Richards, who is said to be sceptical that research is the best way to spend scarce state funds. A panel of prominent researchers from around the state are now finishing a report that they hope will change her mind.

Placating the public in this primarily agricultural state also means retaining many of the service-orientated aspects of the land-grant system. Texas farmers use university researchers as their own private consultants, available to diagnose crop problems and help with new techniques. But administrators realize that strengthening the basic research programme is the best way to attract top scientists, and bovine gene mapping and

Texas research quality improving

Citation impact of Texas science papers in five fields relative to the US averages in those fields.

Source: ISI's Science Indicators Database 1981-91



molecular biology may have trouble outside the university in passing the what's-in-it-for-me test.

The solution, says Shadduck, is to keep an eye on the applied side of the research as well. "We shouldn't let the skepticism of some West Texas rancher prevent us from building a bioscience institute", he says. "But we shouldn't ignore his needs, either."

As with other land-grant universities, Texas A&M is becoming increasingly dependent for survival on attracting funding outside its usual agricultural channels. So far, the school is meeting with success. In the past decade, Texas A&M has risen from eighteenth to seventh among US universities in outside funding, including competitive grants. The Agricultural Experiment Station, a state research and development agency on the Texas A&M campus, has seen its revenues from competitive grants almost double in the past five years, while its state and federal funding stayed flat. The secret, says Robert Merrifield, deputy director of the station, is "taking appropriated funding and leveraging the hell out of it. We use the 'hard' money to groom the researchers to get competitive grants."

In the long run, of course, the university's research quality will determine its funding success. And that means recruiting leading researchers, something at which Texas A&M has had limited success. To many non-Texans, Texas remains a cultural backwater full of cattle and people who wear cowboy hats. College Station conforms to the stereotype. Nothing but top-notch science would persuade most researchers to settle in a town where people routinely drive two hours to get to a good restaurant.

Gerald North, director of the university's new climate systems research programme, was persuaded by Texas A&M's facilities to leave NASA's Goddard Space Flight Center in Greenbelt, Maryland. "I couldn't help but be impressed by the bricks and mortar," he says. "I was convinced that they wanted to make this a first-class university. It still isn't, but I came here because I was convinced that it has tremendous potential."

When he wanted a new supercomputer for climate modelling, he went to the board of regents and reminded them that the University of Texas system had one. "It worked like a charm", he says. He got a new Cray Y-MP for the programme, and another piece of bait to dangle before outside researchers.

New state-of-the-art facilities such as the Crop Biotechnology and the Institute of Bioscience and Technology on its Houston campus will also help. "There's a lot of people I couldn't attract to a vet school in College Station, but I can attract them to Houston," says Shadduck.

The rehabilitation of Texas A&M's sleepy past is going to take more than a few shiny research centres, faculty agree. But their presence improves its chances of enjoying an intellectually rich future.

Christopher Anderson

A&M's plunge into biotechnology has everything — except staff

Houston. Robert Wells sits on the second floor of a brand new and nearly empty 11-storey office building with his hands resting lightly on the unblemished surface of a glowing hardwood conference table. The carpet under his feet is unscuffed. Half of the chairs around the table may never have been sat upon. He and his secretary are not yet on a first-name basis.

Wells, a molecular biologist who moved from the University of Alabama School of Medicine to Texas in 1990, is the director of Texas A&M University's newest experiment in biotechnology — the \$27-million Institute of Biosciences and Technology (IBT) in Houston. Located in the biomedical capital of the state, the institute hopes to provide a link between agricultural and human biotechnology, something that no other centre in the United States has so far accomplished. And putting an agricultural research laboratory in a medical centre is novel, to say the least.

At the moment, Wells, his office staff and his research team are almost the only residents of the IBT, presiding over eerily quiet floors of state-of-the-art transgenic animal facilities and empty conference rooms. Wells spends his time looking for another 400 scientists and support staff to fill the building, an enviable position to be in at a time when his colleagues around the country are struggling to find money to keep the scientists they have. Nevertheless, attracting leading researchers (and their grants) to Texas is not always easy.

One problem is the location. With the Texas biotechnology industry still in its infancy, it takes a good deal of Texan 'can do' spirit to imagine a San Francisco-like biotechnology phenomenon sprouting in the middle of cattle land. Wells explains to

prospective employees that the institute is intended to be part of a new biotechnology industry in Texas, and that the merger of agricultural and human biotechnology builds on the strength of the university research community and the massive Texas Medical Center just down the street.

IBT's research will focus on such subjects as papilloma viruses, molecular parasitology and human and animal genome research. While most private biotechnology funding is going into research on human therapies — where the market is between 10 and 100 times more lucrative than agricultural uses — Wells sees an increased government interest in animal biotechnology. The centre already has two research grants from the US Department of Agriculture, part of what Wells expects will eventually be \$13–\$15 million in outside research funding a year.

This research, in turn, is intended to generate spin-off and start-up biotechnology companies around the Texas Medical Center. The state already has a growing set of small start-up biotechnology companies (see story on page 626), although nothing like the industry that has sprung up around Stanford and the University of California. But with \$1,000 million in new construction (including the IBT) now under way at the medical centre, Texans are hoping that once they build a world-class research centre, the companies and the researchers will come.

By the end of the decade, the Texas Medical Center and its related complexes are expected to employ 50,000 people, a quarter of them involved in basic research. That would be three times larger than the US National Institutes of Health, and one of the biggest biomedical research complexes in the world.

Christopher Anderson

Out, out brief candle...

In March 1989, when the cold fusion sweepstakes began, researchers at Texas A&M University were at the head of the pack. Although results that hinted of excess heat being produced by palladium electrodes in heavy-water electrochemical cells did not stand for long, reports of tritium generation were widely cited as among the most persuasive evidence for some sort of nuclear activity inside palladium electrodes.

The most active proselytizer for the phenomenon was John Bockris, a professor of electrochemistry and longtime friend of Martin Fleischmann, co-inventor with Stanley Pons of the elusive phenomenon. Bockris also had long-standing connections with the Electric Power Research Institute, the research arm of the US electric utility industry, which supplied Texas A&M with funds for cold fusion investigations.

But reports of tritium generation were always sporadic, and during 1990 they subsided under a blizzard of ultimately unproved allegations of deliberate contamination. Some samples of palladium used in the experiments, however, were later found to have contained trace quantities of tritium from the outset.

Claims for cold fusion still pop up from time to time around the world, but not from laboratories in College Station, where neither heat nor tritium has lately been found except in accord with the usual strictures of physics and chemistry.

David Lindley

Energy researchers fight for their survival

Fort Worth. Energy research in Texas is at a crossroads. Gone are the good old days when oil and gas fuelled the state economy and provided steady jobs for geologists and other researchers. In fact, the state's business base has diversified and energy's economic contribution has fallen so much that the Texas Employment Commission has reclassified oil and gas as a 'declining industry.'

Most large oil companies have made deep cuts in their Texas work forces in reaction to falling oil and natural gas prices and declining demand for crude. In recent years, thousands of scientists in the laboratory received pink slips along with the roughnecks working on the rigs.

Geophysical scientists involved in exploration have fared even worse. With new discoveries in the lower 48 states increasingly rare, US companies looking for new oil are most likely to do so abroad, offshore or in Alaska.

As a result, the private sector research laboratories that once fed the nation's largest oil-producing state are in transition. They are shrinking in size, and their focus is changing. They are trying to get more oil from existing fields by using sophisticated computer-based technologies such as imaging and simulation, and they hope that the same technology will also allow them to do it faster, cheaper and more efficiently.

Because oil companies have less money to spend, they are also trying to make every research dollar count — through joint ventures and industry consortia as well as joint programmes with university research centres. "The industry needs to be more intelligent about how to spend its research dollars", says Michael Wiley, a vice president at Atlantic Richfield Co. and manager of ARCO's exploration and technology operation in Plano, a Dallas suburb.

Oil & Gas Journal, an energy industry trade publication, estimates that US companies will spend \$32,500 million on domestic energy projects this year, compared with \$34,600 million in 1991. The journal's annual survey found that exploration and production budgets, a subset that accounts for most spending on research, will reach a five-year low this year.

Although universities in Texas operate a number of energy research centres, private industry has historically driven the state's oil and gas research. At the same time, however, those companies finance mostly applied research. They leave basic research to universities, national research laboratories and industry consortia. Even though

research in the private sector is proprietary in nature, much of it eventually becomes public through technology transfer.

That is why the latest research figures worry industry officials such as Bill Fisher, director of the Bureau of Economic Geology at the University of Texas at Austin. The bureau is the state's main energy research arm, housing the State Geological Survey and emphasizing applied research in oil and gas. "We're losing a lot of very experienced talent in the industry," Fisher laments.

Some of the displaced geologists and petroleum engineers are finding jobs in related fields that also involve geophysical

oil and gas a vital industry in Texas. Their contribution to the gross state product has declined in the past 30 years from 20 per cent to about 15 per cent. New technology could prevent it from slipping further, according to the Bureau of Economic Geology.

ARCO's exploration and technology research group, which supports all of ARCO's operations, has long operated an enhanced recovery programme using such technologies as advanced simulation. "We're probably doing more in integrating that technology more closely with geoscientists than in the past," Wiley said. "Everything we do today is cross-disciplinary."

Medium-sized independent companies and universities are likely to play a larger role in researching and developing such advanced technologies. Many companies have been funding joint research programs at universities for years, and the idea is expected to become more popular.

Research centres such as the Bureau of Economic Geology and trade groups such as the Texas Independent Producers & Royalty Owners have joined to help transfer more technology to smaller companies that lack money to carry out research. The two are teaching smaller companies how to use computer-based technology to remain competitive — or at least to survive the latest wave of contraction.

Industry would also like the state government to fund more energy research and find new markets for energy-related products. The state now spends only about \$5 million annually on oil and gas-related research. But an increase is not likely until the state's budget problems ease.

Fisher says annual research expenditures on Texas energy resources must rise from the current \$50 million to \$300 million annually in five years to take advantage of the state's remaining oil and gas resources. That would include an increase in state spending to \$30 million annually and a jump in federal spending from \$10 million to \$200 million annually with a quarter of the money spent in Texas. Raising federal spending to that level would still equal only 0.3 per cent of the current trade deficit from imported oil, Fisher estimates.

Although some energy industry executives say such generous outlays are just wishful thinking, Fisher and other state energy officials say that a tax credit incentive, now being debated, could finance much of the increase. The outcome of that campaign could help to determine the shape and make-up of the Texas economy well into the next century. Recognizing the industry's need for continued technology advances, says Fisher, would allow oil and gas to remain a vital part of the New Texas.

Kathryn Jones

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Top, ARCO geologists in Houston study maps of oil fields; left, a computerized image of an underground formation.

science—hydrology or environmental protection against underground waste storage tanks, for example. Others have taken jobs abroad, left the volatile industry or continue to look for work.

Historically, oil and gas research in Texas has concentrated on exploration to find and develop new fields. But the greatest potential for future oil production in Texas may be recovering more oil from known fields.

The bureau estimates that existing Texas fields contain 170,000 million barrels of oil. Industry has recovered only 35 per cent, and sent most of it to refineries. About a third of the remaining oil is accessible by conventional means and resides mostly in geologically complex reservoirs in the so-called Permian Basin of West Texas.

Removing this oil depends on advanced geophysical detection technology and advanced geophysical modelling, Fisher says. Recovery of the rest will require advanced extraction technology. Such technology could double the amount of oil recovered in the state from existing fields, Fisher reported last year, and nearly double the existing volume of natural gas.

Using such technology could help keep

Photo by Lois Gervais, Atlantic Richfield

Space science in Texas means basic research as well as support for manned flights

Houston. In Texas, space usually means the Johnson Space Center and its nearly 17,000 engineers and technicians. But behind this army of designers of the manned US space programme is a network of smaller institutions that are as focused on basic research as Johnson is on engineering.

This diverse network may turn out to be key to the future of space research in Texas. At the moment, however, researchers are working under the threat of static (or worse) funding of the US National Aeronautics and Space Administration (NASA), whose proposed budget of \$15,000 million in 1993 is hostage to the recession and the federal budget deficit. It may be years before Texas again sees anything like the heady growth of the 1960s, when the Apollo programme put Texas space science on the map virtually overnight.

NASA's recent success at satellite retrieval may have helped to build the case for a continued human presence in space, but the programme's future is still tied to the space station, a frequently redesigned project with more than its share of political opposition. NASA's most memorable recent science successes — the data from the Cosmic Background Explorer that confirmed the Big Bang theory and the Neptune flyby of the Voyager probe — were both robotic missions, and the continued high cost of the manned programme has led the agency to look more favourably upon smaller, cheaper unmanned missions (see *Nature* 357, 184; 1992).

With direct NASA funding becoming an increasingly rare commodity, Texas's small university laboratories and research institutions are finding that the best way to stay in the game is to reach out and form collaborations with institutions outside the state.

One such facility is the Lunar and Planetary Institute (LPI), created in 1968 to distribute lunar material from the Apollo landings to scientists outside NASA and to serve as a clearinghouse for planetary information. Although it receives funding from NASA and works with the Johnson centre, LPI also enjoys a close relationship with scientists at the Jet Propulsion Laboratory in Pasadena, California, who design many of the US robotic planetary missions. This broad network allows LPI to benefit from both the manned and unmanned programmes, without depending too much on either.

The LPI has also developed small-scale

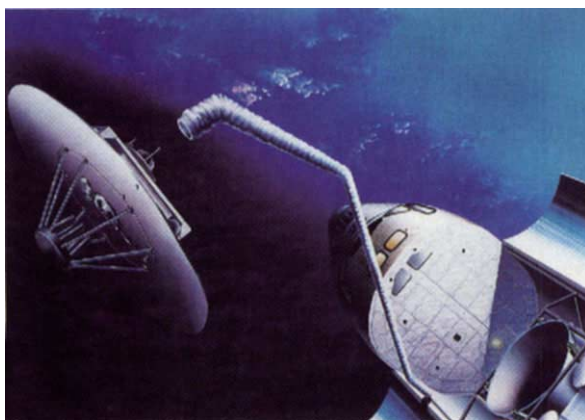
collaborations with private industry. Its scientists are working with petroleum companies to analyse proprietary underground data in the hope of generating a better understanding of the cratering history of the Earth. LPI scientists also are studying core samples taken by Pemex, the Mexican national oil

from the Magellan mission to map the surface of Venus and from the Hubble Space Telescope. When the Galileo probe arrives at Jupiter in 1996, LPI scientists hope to support that mission's ambitious planetary objectives too, assuming that JPL engineers have either jostled Galileo's defective antenna into position or figured out

another way to capture the data from the spacecraft's recorders.

Researchers at LPI's parent organization, the Universities Space Research Association, are concentrating on some of the life science aspects of space research. They are involved in some Johnson shuttle experiments, as well as providing support to an assortment of international projects such as the joint US-Russian life sciences working group. But funding and employment in this area have never been large, and researchers expect only a modest growth in opportunities if and when the space station Freedom becomes operational.

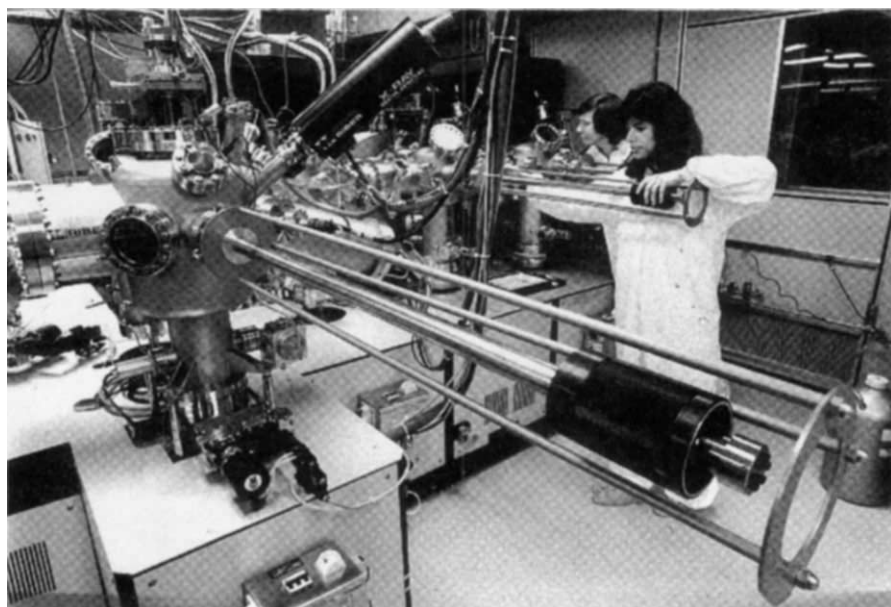
At Texas universities, where several space science programmes also got their start with the Apollo effort, the future also appears to lie with outside collaborations and small projects. The first space science department in the world was established in 1963 at Rice University in Houston. Now, however, its faculty is more involved in projects carried out by out-of-state centres such as JPL, Goddard



Drawing shows wake shield shortly after it is released by the orbiting space shuttle.

company, from an impact crater basin in Mexico's Yucatan peninsula. Some argue that the crater was caused by an asteroid that triggered the extinction of the dinosaurs some 65 million years ago.

At present, some of the 60 LPI scientists and staff are digesting a steady flow of data



Gallium-arsenide wafers being processed at Houston's space vacuum epitaxy centre.

Space Center in Maryland and Marshall Space Flight Center in Alabama. Rice scientists are concentrated in astronomy and plasma, with roles in the Hubble mission, the Global Geothermal Satellite project and joint missions with the Europeans and Japanese.

As a small centre, the Rice space science programme has found itself vulnerable to the fluctuations of the US space programme. Eight million dollars of annual funding for the operation of three NASA solar plasma and physics missions has shrunk to almost nothing over the past two years as NASA decommissioned the probes. Analysis of the old data is all that remains of the three solar programmes at Rice, but the university hopes to become involved in future missions.

Another Texas university space science group that must avoid placing too many of its eggs in one basket is the University of Houston's Space Vacuum Epitaxy Laboratory. Funded since 1987 through NASA's Office of Commercial Programs, the laboratory has been at the forefront of work on novel vacuum-growth technologies both on Earth and in space, in particular thin-film materials. In November 1993, the laboratory hopes to launch from the bay of the space shuttle a stainless steel wake shield, a disk 12 feet in diameter, that will trail the shuttle. It is intended to create a nearly perfect vacuum in which to manufacture materials for better computer microchips.

A potential bonanza if it succeeds, the experiment has only a small foreign contribution (from the Canadians) and must demonstrate its commercial viability quickly. Because the technology is in the public domain, however, it is hard to keep secret: the Japanese, for instance, plan to fly a similar mission in 1994 on the basis of information they have obtained from the Houston effort. And its first launch is critical: researchers must demonstrate to NASA that the wide-shield technology is feasible to win support for the remaining four launches in the series.

Despite the hurdles, Texans consider space science to be an important element in the state's plans for technological growth in the 1990s. Although the Johnson centre and the surrounding institutions expect to grow little throughout the decade, state officials hope that a new regional technology transfer centre in Houston, funded by NASA, will serve as a broker of space services between the government and industry. At the same time, they expect the centre to attract to Texas other federal agencies — including the Departments of Defense, Energy and Transportation — with money to spend on space science. And if the space station makes it through the political grinder, Texas's focus on manned space may actually start to look like an advantage.

Alcestis Oberg

San Antonio aims for spot on high-technology map

San Antonio. When most people think of San Antonio, the first thing that springs to mind is the Alamo, the Franciscan mission where, in 1836, Texans were besieged and massacred by Mexican troops. Today, San Antonio is a culturally diverse city, with strong social and economic ties to Mexico. And it is home to many distinguished military and civilian research facilities. Included among these are the Armstrong Laboratory at Brooks Air Force Base, Brooke Army Medical Center, Wilford Hall Medical Center at Lackland Air Base, the University of Texas Health Science Center at San Antonio, the University of Texas at San Antonio, the Cancer Therapy and Research Center of South Texas, the Southwest Foundation for Biomedical Research and the Southwest Research Institute. The last two were founded by a local industrialist, the late Thomas Slick, Jr, who studied biology at Yale University and who died in an airplane crash in 1962 at the age of 46.

With all this academic and clinical research, it may seem surprising that there are only four pharmaceutical and five biotechnology companies in the San Antonio area. But, whereas Houston has three or four venture capital firms willing to seed new biomedical ventures, San Antonio lacks a venture capital community of its own and has failed so far to attract the attention of out-of-state investors.

Despite the odds, the Texas Research and Technology Foundation (TRTF) has been promoting economic development in south and central Texas since 1984. Funded largely through philanthropic gifts and with no state or federal government support, TRTF owns and manages the 1,500-acre Texas Research Park situated in rolling hill country about 20 miles west of the city. The land was donated to TRTF in 1986 by Tom Pawell, president of the Concord Oil Company.

The first major research facility to be built within the park is the University of Texas Health Science Center's Institute of Biotechnology, funded by a \$15-million grant from the family of industrialist and presidential hopeful H. Ross Perot. This two-storey, 58,000-square-foot building is expected some day to house about 110 scientists, post-docs and support staff.

Just a stone's throw away, construction is under way on two additional and complementary research facilities — the Cancer Therapy and Research Center's Institute for Drug Development and the national headquarters of the Southwest Oncology Group (the largest clinical trials organization in the United States). Due to

be completed by the autumn, the institute is headed by Daniel Von Hoff, one of only six investigators selected by the National Institutes of Health's cancer institute to direct phase I clinical trials. Its mission is to shorten the time required to develop new anticancer agents.

Groundbreaking ceremonies have yet to be held on the new campus for the Southwest Foundation for Biomedical Research, which plans to relocate to the park from a site closer to town. The foundation maintains the world's largest colony of baboons for use in medical research and conducts research on organ transplantation, cardiovascular diseases, AIDS, premature birth and behavioural medicine.

As well as recruiting companies to the park, TRTF is also helping small start-up companies and individual researchers by subsidizing laboratory and office space. Researchers are using these 'incubator' facilities to test the commercial viability of their research projects before raising venture capital. The current crop of researchers are expected to outgrow their incubator facilities within three years, at which time they will have the chance to lease or purchase land or facilities within the park. In addition to the incubator facilities, TRTF hopes that the manufacturing facility being built on site to meet US Food and Drug Administration specifications will attract companies to the site.

Two years ago, TRTF spun off a for-profit venture capital company called the VenTex Group, Inc. to make limited funding available to new companies. VenTex raises and manages seed and growth capital for high-technology companies, and invests in the biomedical, environmental, computer and material sciences. Charles McLaughlin, president of VenTex, says the company is willing to back research projects at a very early stage and will provide up to \$250,000. So far, VenTex has made five investments.

It is too early to judge whether TRTF will succeed in its efforts to attract a healthy mix of academic institutions, start-ups and more established high-technology companies to the park and stimulate the local economy. But there are signs that its job will not be easy. Three years after ground was broken for the Institute of Biotechnology, more than 75 per cent of the park remains undeveloped. And so far TRTF has failed to entice an existing biotechnology or pharmaceutical company to the park.

Diane Gershon

Thanks to Barbara Ann Bukowski for her helpful information on bioscience industries in the Southwest.

The objective case for conservation

Not before time, people are beginning to wrestle with the concept of biodiversity not simply from academic inclination, but so as to forge a practical yardstick for conservation.

Two neologisms beginning with 'b' have crept into general use in the past year or so: 'buckyballs' and 'biodiversity'. What the former lacks in elegance, the latter lacks in definition. Which is odd, considering that 152 world leaders have recently signed their names to a pledge to protect biodiversity, whatever it may be. But that circumstance was enough to lift the hearts of participants at a meeting* in London last week (planned long before the Rio summit) designed to grapple with the task of defining biodiversity, how to measure it, and how then to conserve it. ('Biodiversicist' may well be among next year's crop of unwieldy neologisms.)

But measuring biodiversity means more than making a list of species. Species are in any case disappearing more quickly than they can be formally named, so that a comprehensive inventory would arguably be more useful to a recording angel than to conservationists working to save what are left. Formal lists are out — "quick and dirty" lists (in the words of Robert M. May of the University of Oxford) are in. But such lists must include details of habitat, range, conservation status and phylogenetic relationship as well as a taxonomic label.

Inventories with these features are already beginning to be compiled and used, in particular to identify biodiverse 'hotspots' — circumstances, topographical or otherwise, notable for their diversity. It matters little that some of this information may be missing, because cross-referencing of what information there is should be enough to inform provisional decisions about conservation priorities. Time, money and manpower are in any case brakes on what can be done.

Biodiversity, then, is an epiphenomenon, an expression of the interaction of a variety of biotic and abiotic factors. Redundancy in the system ensures that the conservation of biodiversity remains valid as an ideal, even when many (or most) natural habitats have disappeared.

Accordingly, biodiversity embraces far more than taxonomy, although access to biological collections and taxonomic expertise is still central. This is reflected by the diversity of approaches now used to measure biodiversity, which can be quantified by the span of phylogenetic relatedness (the more, the more diverse), using cladistic or cytogenetic language; by the patchiness of

habitats and biogeographical ranges; by the abundances of key 'indicator' species; by the identification of particularly important ecological interactions; or combinations of some or all these approaches.

Inevitably, many of the yardsticks being canvassed are mathematical in character, not entirely in tune with the popular image of the conservationist as a person with a big heart and, if a man, with a chunky hand-knitted sweater and a beard. The immediate task is to ground emotion in objective measures. The goal, throughout, is the same: how to conserve the greatest amount of biodiversity in the smallest area, and so (presumably) at the lowest possible cost.

But conservation policies, however well-rooted in science, can work only if they are intelligible within the diverse cultural contexts of those required to follow and to enforce them. To parrot Marx, conservation policy is both idiosyncratic and eclectic. It may follow that the meaning of biodiversity itself is coloured by cultural referents. That a single word can encompass so much may be sufficient excuse for both its awkwardness and its vagueness.

But why conserve biodiversity at all? It is difficult to argue for conservation if biodiversity is represented only by a list of species, some (but not all) of which may be technically endangered. Of course, there may be moral or aesthetic grounds for the conservation of individual species, but even these may differ between cultures (witness the annual conflict, about to be resumed in the International Whaling Commission, between whaling nations and others). Similarly, when people from rich countries decry the loss of habitat elsewhere, they may find their arguments cut no ice with those to whom habitat loss may mean the key to short-term economic survival.

Although the loss of the African elephant or the blue whale would be sad, there is no particular reason to spare these species in particular if they are just two among many millions. But to argue that the blanket preservation of millions of species might safeguard the few that, at some unknown future date, might prove useful in agriculture or medicine sounds like special pleading. Such possibilities are by-products of some elements of biodiversity, and offer no good reasons for conserving biodiversity in totality.

Far more convincing arguments stem from the concept of biodiversity as the sum of the interactions between species, rather than a list of the species themselves. More than this, these arguments follow from this opera-

tional concept quite naturally, and do not have to be tacked on as afterthoughts to justify the effort.

That biodiversity defines a working system rather than a list of parts means that the conservation of as many parts of it as possible allows the system as a whole to retain flexibility (which can be a hedge against change) as well as preserving diversity for future scientific study or commercial exploitation. Importantly, these changes embrace far more than natural variations in climate.

There are, of course, some who argue that conservation is futile, given that evolution and extinction are natural consequences of climate change, irrespective of human interference. But that argument overlooks the vested interest of human populations in keeping things very much as they are, which implies that conservation has a natural place in economic strategy. 'Conservation', after all, has the same etymological roots as 'conservative' (with a small 'c').

It follows from this that biodiversity plays a real part in the material well-being of people. How can this be? Few will be able to identify — let alone quantify — any coarsening in their quality of life as a result of the extinction of this species or that. But it is certainly the case that the physical circumstances that allow us to live and breathe on the Earth are in no small measure the products of biodiversity, in the sense that biodiversity is a system rather than a list. Biodiverse interactions over billions of years have ensured that we have an atmosphere to breathe, a soil to till and fresh water to drink (and that we ourselves are adapted to the use of these materials, not to others not in existence). The integrity of this system as a whole transcends the extinction of particular species, even if some species may be more important than others for the coherence of the overall fabric.

The pollution of the atmosphere, the denudation of soil and the diminution of supplies of fresh water in many parts of the world could be seen, in part, as consequences of the loss of biodiversity. As longtime environmental campaigner David Bellamy put it at the start of last week's meeting, the effects are already being felt. Economic hardship and civil strife may be consequences, in part, of the neglect of conservation as well as its continuing causes. Perhaps governments' acknowledgement of that explains why the biodiversity treaty gathered so many signatures at Rio. The worry is that they may not have been fully aware of what the concept means.

Henry Gee

*Systematics & Conservation Evaluation, Natural History Museum, London, 17–19 June 1992.

Hard-pressed molecular clouds

Mark Morris

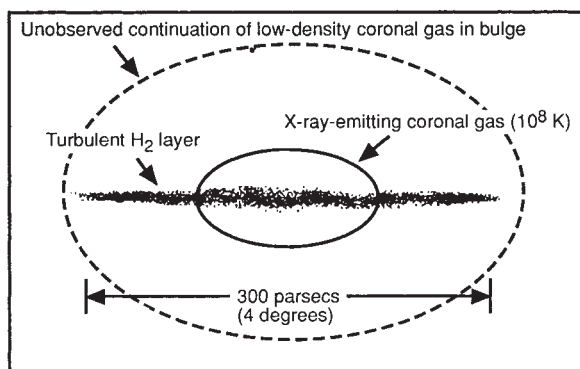
JUST about everything — stars, gas, magnetic fields and related activity — gets concentrated towards the depths of the central gravitational potential well of our Galaxy. So it is not particularly surprising that the average pressure of the interstellar medium is higher there than in the distant disk of the Galaxy, where we reside. Indeed, observations have long shown that the gas there is substantially warmer and denser than anywhere else in the Galaxy¹. However, the arguments presented by Spergel and Blitz on page 665 of this issue², that the interstellar pressure within a few hundred parsecs of the nucleus is a full two and a half orders of magnitude larger than that of the solar neighbourhood, threaten to force a reassessment of our notions of the flow of thermal and dynamical energy through the Galactic Centre.

The interstellar medium within a few hundred parsecs of the galactic nucleus (we are 8 kiloparsecs from the centre) is dominated by two markedly contrasting phases (see figure): a planar distribution of molecular clouds with a total mass of about a hundred million solar masses, representing a far greater concentration of cool, dense clouds than is found anywhere else in the Galaxy; and an elliptical cloud of hot, coronal gas which is five or six orders of magnitude less dense than the molecular clouds, but about six orders of magnitude hotter. The coronal phase has been directly observed as an extended source of X-rays, both in the continuum^{3,4}, and in the 6.7 keV line of highly-ionized, helium-like iron⁵, which is characteristic of a 10^8 K plasma.

The X-ray results had already implied that the coronal phase has a high pressure, but Spergel and Blitz are apparently the first to consider the effect of such pressure on the molecular cloud layer. We have known for about two decades that the widths of molecular spectral lines observed towards Galactic Centre clouds are much larger than in disk clouds¹. Interpreted in terms of turbulent velocities, as is done by Spergel and Blitz, such linewidths imply a turbulent pressure large enough to be in equilibrium with the high pressure of the surrounding coronal medium. It is also large enough to resist the self-gravity of such clouds and therefore either to impede star formation altogether, or to

allow only low-mass stars to form. This may explain why there is not much evidence of star formation in the Galactic Centre region by comparison with clouds in the galactic disk.

Several important questions are raised by the conclusion that the interstellar pressure is so large in the Galactic Centre. The first is, what is the tremendous source of energy required to maintain the high pressure of the interstellar gas?



The reservoir of molecular gas (mostly H_2) in the Galactic Centre is arrayed along the plane of the Galaxy and extends a few degrees (150 parsecs) in both directions away from the nucleus. We see it from an almost edge-on perspective (although it may be somewhat tilted), so it appears to the radiotelescopes which are used to map it as a relatively thin line of emitting material. The hot, coronal gas, on the other hand, has been observed with X-ray telescopes to be about half as extended along the galactic plane, but much more broadly distributed into the stellar bulge above the plane. The scale of the central stellar bulge, about 1 kiloparsec, is much larger than either of the observed interstellar constituents, although the coronal gas may extend out into the entire bulge with decreasing (and hence unobservable) density. The density of stars in the central bulge (not included in the figure) increases as $(\text{radius})^{-1.8}$ all the way into the inner parsec.

If supernovae are responsible, as many as one supernova every 100 years may be needed in the inner 150 parsecs (ref. 5), comparable with the rate that is often assumed for the entire Galaxy. Alternatively, a single explosion releasing 10^{54} ergs of energy within the past 10^5 years could supply this energy⁵, although other evidence for such an event is lacking. Other sources should clearly be considered, such as the gravitational potential energy released by gas accreting into the central region.

Second, what is the scale height of the coronal gas layer? If it is only as large as the observed size of the X-ray emitting regions, then Spergel and Blitz note that the high gas temperature would drive a galactic wind. The mass flux in such a wind would be distressingly large, approaching a few tenths of a solar mass every year. Over the lifetime of the

Galaxy, such a rate of mass loss would imply that the Galactic Centre has lost a quantity of mass comparable to the total mass now present there in the form of both stars and gas. On the other hand, if the scale height were large enough (1 kiloparsec) to permit a hydrostatic description of the coronal gas, then the entire galactic bulge should be luminous in X-rays. Do current observations rule out such a high-luminosity source in the central bulge of the Milky Way or in other nearby normal galaxies?

Lastly, can the turbulence which presumably supplies the molecular-cloud pressure be maintained? In a dense cloud, turbulence is dissipated on a relatively short timescale, so it must constantly be replenished. What keeps the clouds so effectively stirred up? The interstellar magnetic field is another source of pressure, and perhaps a very important one. At various places within the inner 100 parsecs, the magnetic field strength has been estimated to be of the order of a milligauss⁶, much stronger than is required for pressure equilibrium with the turbulent cloud pressures and the thermal pressure of the coronal phase. Perhaps there is a strong and pervasive magnetic field which dominates the pressure of all phases.

If so, the cloud turbulence might be reinterpreted as magnetohydrodynamic waves there, as it has for clouds in the disk⁷. This might circumvent potential problems with the short timescales for the dissipation of turbulent energy, as the rate of dissipation of the wave energy is probably not as rapid. A strong magnetic field would also confine the coronal gas, and if the field has an organized poloidal geometry, as some evidence may indicate, then it provides a chimney for thermal wind to rise out of, producing, perhaps, the vertical radio structures seen in the nuclear region of a number of spiral galaxies, including our own⁸. □

Mark Morris is in the Department of Astronomy, University of California, Los Angeles, California 90024, USA.

1. Güsten, R. in *The Center of the Galaxy* (ed. Morris, M.) 89–106 (Kluwer, Dordrecht, 1989).
2. Spergel, D. N. & Blitz, L. *Nature* **357**, 665–667 (1992).
3. Skinner, G. K. et al. *Nature* **330**, 544–547 (1987).
4. Kawai, N. et al. *Astrophys. J.* **330**, 130–141 (1988).
5. Yamauchi, S. et al. *Astrophys. J.* **365**, 532–538 (1990).
6. Morris, M. in *Galactic and Extragalactic Magnetic Fields* (eds Beck, R., Kronberg, P. & Wielebinski, R.) 361–367 (Kluwer, Dordrecht, 1990).
7. Arons, J. & Max, C. E. *Astrophys. J.* **196**, L77–L80 (1975).
8. Hummel, E., van Gorkom, J. H. & Kotanyi, C. G. *Astrophys. J.* **267**, L5–L9 (1983).

Vascular pipe dreams

Alan R. Hemsley

THE definitive evidence of vascular tissue in the axes of *Cooksonia pertoni*, presented by Edwards *et al.* on page 683 of this issue¹, finally confirms the belief held by many palaeobotanists — *Cooksonia* was indeed a vascular plant. The discovery strengthens the position of *Cooksonia* as an ancestral vascular plant (Fig. 1) and supports generally held ideas regarding the evolution of early land plants.

Cooksonia pertoni is of upper Silurian

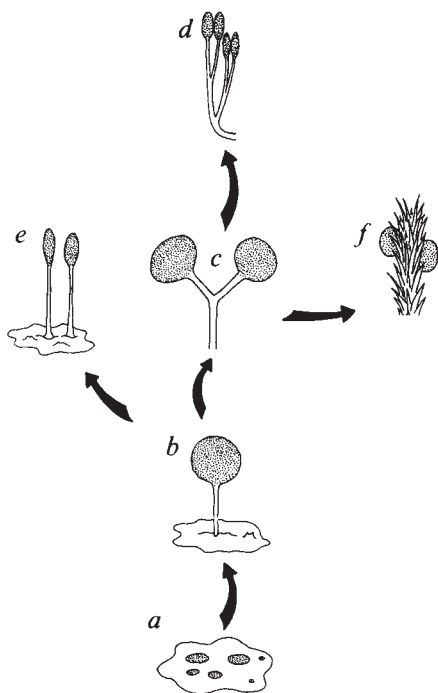


FIG. 1 A possible phylogeny of land plants illustrating the important position held by *Cooksonia*. a, A thalloid alga bearing simple sessile sporangia (stippled). b, A thalloid gametophyte supporting a simple sporophyte consisting mostly of sporangium. c, *Cooksonia*, a branched sporophyte. d, An early Devonian vascular plant showing dichotomous branching and numerous sporangia. e, A simple bryophyte, possibly derived from b. The lycophytes (f) appeared at about the same time as *Cooksonia* and may be derived from a similar ancestor.

to lower Devonian age — that is, is some 400 million years old — and was first described in 1937 (ref. 2). It was a regularly branched plant only a few centimetres in height, bearing rounded spore-bearing structures (sporangia) at the end of upright axes. The apparent simplicity of this structure gives the plant a kind of 'totipotency' in terms of its potential for the subsequent evolutionary development (in Devonian vascular plants) of the more complex branching patterns of the sporophyte. The discov-

ery of thickened tubes in *Cooksonia* therefore provides the basis for the elaboration of not just the morphological traits, but also the water-conducting system. It is the dual role of water transportation and support performed by vascular elements within sporophytes that has permitted the increase in size beyond the diminutive *Cooksonia* to the arborescent state seen in a number of plant groups later in the Devonian.

Despite the earlier absence of evidence regarding its vascular nature, it has for some years been known that *Cooksonia* possessed other crucial characteristics generally associated with land plants³: a cuticular covering, stomata and thick-walled spores^{4,5}. But such a combination also occurs within bryophyte sporophytes, and so the status of *Cooksonia* in terms of its relationship with the two main groups of land plants, the vascular and nonvascular, remained unclear.

Investigation of slightly younger Devonian material^{6,7} has begun to reveal the distribution of the four main types of water-conducting tissues found in early land plants — simple elongate tubes; S-type cells⁶ with an inner microporate layer covering the thickening within the tube; G-type cells⁶ with simple annular thickenings; and tracheids with scalariform pitting (Fig. 2). Simple tubes and S-type elements are present in the axes of many of the early Devonian plants but not within the lycopods, where the G type predominates. Tracheids are not encountered unequivocally until the late early to middle Devonian. Significantly, some living liverwort gametophytes, such as *Takakia* (considered primitive within this group⁸), possess the S-type element.

The vascular elements found in *Cooksonia* do not appear to have an inner microporate layer, and were probably of the G type. However, the morphology of the plant bears less resemblance to the lycopods and more to the other contemporary groups where S type and unornamented tubes occur. It seems most reasonable that tracheids with scalariform pitting should emerge ultimately from the G-type element, this being a satisfactory hypothesis for the occurrence of tracheids in lycopods but less adequate in explaining their presence in the presumed descendants of the S-type plants. The occurrence of G-type elements in *Cooksonia* offers the possibility that tracheid-like elements were present relatively early amongst some non-lycopod groups. Other groups possessing S-type elements (including ancestral

RÉSUMÉ

Gangliosidomulin?

GANGLIOSIDES are components of membranes, especially in nerve cells, and play a still ill-defined part in proliferation and differentiation of cells. H. Higashi *et al.* (*J. Biol. Chem.* **267**, 9831–9838; 1992) have looked for and found a ganglioside-binding protein from brain cytosol, which, on examination, turned out to be nothing other than calmodulin. Gangliosides are known in certain cases to affect the activities of enzymes that are themselves controlled by calmodulin. A second paper in the same issue, by H. Higashi and T. Tamagata (9839–9843), reports on such a system in which the gangliosides bind to both the enzyme and its regulator, calmodulin. A scheme, involving the displacement of gangliosides from the enzyme by the calmodulin, is elaborated.

Down the tube

THE axion, proposed to explain why neutrons do not have the large electric dipole otherwise expected, is supposedly a nearly massless particle (like the neutrino) with no angular momentum (unlike any other elementary particle). They should be produced copiously in the Sun and other stars — if they exist. The problem is finding a suitable telescope to pick them up; a long length of vacuum tubing surrounded by magnetic fields is what is called for, and can be found in most particle accelerators. But these are hardly manoeuvrable. However two straight sections of the LEP storage ring at CERN point just 1.4° from the midsummer sunrise position, note F. Hoogeveen and R. G. Stuart (CERN preprint TG.6457/92) — as future archaeoastronomers may spot, they add mischievously — and twice a year (in summer and winter) the Sun shines straight through them for about 20 seconds. Could LEP enjoy a transitory change from particle factory to astronomical telescope?

Model prospects

IN a report to appear in next week's *Science* (3 July), M. B. Agy and colleagues show that the pigtail macaque is susceptible to the type-1 human immunodeficiency virus. The work stems from the need to find a different animal model for research into HIV-1 — chimpanzees can be infected with the virus (although they do not develop AIDS) but are an endangered species. Pigtail macaques are not, and the eight individuals inoculated by Agy *et al.* all showed heavy and persistent infection with the two virus strains used. The species may well turn out to be at least as appropriate as the chimpanzee as a model for examining the initial course of infection with HIV-1 and evaluating potential vaccines.

Raising the roof of the world



HIMALAYAN mountains such as Everest, pictured here from Tibet, may have been raised higher in the late Cenozoic by an increase in erosion, according to Douglas Burbank (page 680). Perhaps paradoxical at face value, this conclusion is based on the pattern of river flow and sediment accumulation in the Indus-Ganges foreland basin of Pakistan and India, and it supports the hypothesis (P. Molnar and P. England *Nature* 346, 29–34; 1990) that mountain uplift can be climatically driven. The argument is as follows. Climate change, such as the cooling that took place in the late Cenozoic (the past 15 million years), might increase the rate of erosion by causing changes in precipitation and storminess. Removal of surface material then produces isostatic uplift of the range as a whole.

liverworts) may represent sister groups to a lineage where G-type elements eventually gave rise to tracheids with scalariform pitting.

That the nexus of vascular elements

reproductive morphology and in propagule type, much of which may have stemmed from a basic 'looseness' in the underlying developmental processes. Indeed it would be surprising if the early

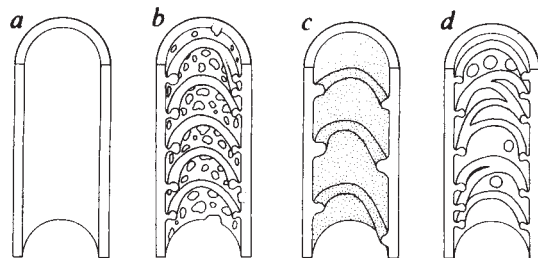


FIG. 2 The four types of vascular element described in the text. *a*, A simple tube element. *b*, The G-type element with a perforate wall and thickenings. *c*, The S-type element with a microporate inner layer covering the thickenings and wall. *d*, A tracheid with scalariform thickenings. (*b* and *c* after ref. 6.)

present in the early Devonian is not wholly compatible with our concepts of the major plant groups is perhaps not surprising. All too often in these early land plants, those features which are recognized today as being characteristic of a particular evolutionary line or clade are distributed on a 'mix and match' basis. It has been noted⁷ that differences between G-type, S-type and scalariform elements may have a developmental basis, and it is perhaps possible that changes in the degree of ornamentation or production of any of the components (thickenings, cellulosic layers, microporate structures) could give rise to any of these types of conducting element. The factor governing the outcome of the developmental process is not the ability to produce only one type of conducting element, but the degree to which a particular developmental process has become established and predominant.

Many features of the early land plants (including *Cooksonia*) point to there having been a measure of plasticity and flexibility, as seen in basic vegetative and

land plants, which had undertaken one of the most dramatic changes of habitat experienced by plants, were not to have had a considerable flexibility in their requirements and responses. That phenomenon is perhaps being illustrated in the variety of conducting elements now known to occur in these early land invaders. □

Alan R. Hemsley is in the Departments of Biology and Geology, Royal Holloway and Bedford New College, University of London, Egham Hill, Egham, Surrey TW20 OEX, UK.

1. Edwards, D., Davies, K. L. & Axe, L. *Nature* **357**, 683–685 (1992).
2. Lang, W. H. *Phil. Trans. R. Soc. B* **227**, 245–291 (1937).
3. Chaloner, W. G. in *Proc. XIV. Int. Bot. Congr.* (eds Greuter, W. & Zimmer, B.) 301–316 (Taanus, Königstein, 1988).
4. Edwards, D., Fanning, U. & Richardson, J. B. *Nature* **323**, 438–440 (1986).
5. Fanning, U., Richardson, J. B. & Edwards, D. *Evol. Trends Pl.* **2**, 13–24 (1988).
6. Kenrick, P., Edwards, D. & Dales, R. C. *Palaeontology* **34**, 751–766 (1991).
7. Kenrick, P. & Crane, P. R. *Bot. Gaz.* **152**, 335 (1991).
8. Schuster, R. M. in *Bryophyte Systematics* (eds Clarke, G. C. S. & Duckett, J. G.) 41–82 (Academic, London, 1979).

Rather than focusing on records of uplift and denudation in the mountains themselves, Burbank looked for sediment removed by increased erosion. He considers two extreme cases: one in which uplift was not caused by erosional unloading at all but by tectonic forces, specifically by accelerated underthrusting of the range by the Indian plate, and the other in which only erosion played a role. The first produces subsidence and deposition of sediments, and therefore steep slopes close to the foot of the range. Erosional unloading, however, increases the sediment influx into the basin, creating gentle slopes far into the foreland. In this case rivers can run parallel to the mountain range only far from it.

In principle, therefore, the distinction between the two cases can be made simply by looking at a map of the Ganges and Indus hinterlands. The two rivers tell different stories, however: the Indus is fed by tributaries running transverse to the Himalayan range, as predicted by an erosional scenario, whereas the Ganges system is longitudinal, suggesting the predominance of thrust loading. But the depositional patterns have changed through time. The Ganges system probably switched from transverse to longitudinal around 4 million years ago, post-dating by a considerable margin the principal uplift of the Himalayas. So erosional unloading seems to be the better candidate; but whether this can be attributed to climate change remains to be seen.

P.B.

Homing in on the homeobox

Guy Riddihough

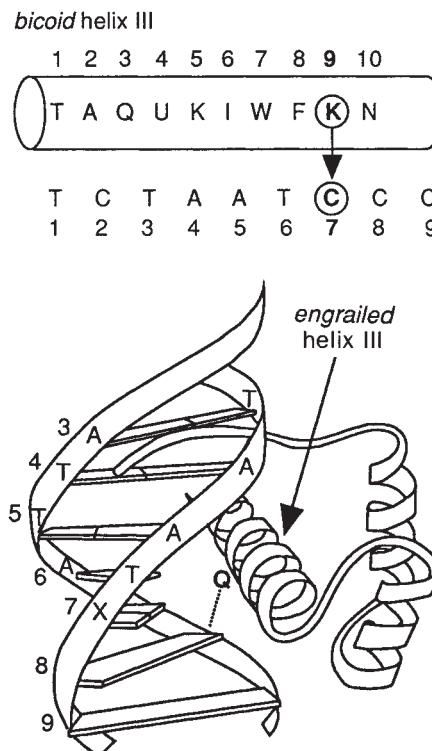
THE homeobox is the Rosetta Stone of developmental biology. Study of this highly conserved DNA sequence, which encodes a homeodomain of 60 amino acids, is opening the way into understanding how the body plans of various animal phyla are laid down. But what is the mechanism of action of the homeodomain, at the levels of interaction with DNA, of determining cell and tissue fate, and of evolution of organismal form? This question, in all its rich complexity, was the grist for discussion at a meeting* held last month.

The homeodomain folds into three α -helices (I, II and III), the second two adopting a helix-turn-helix conformation characteristic of one of the main classes of DNA-binding protein; indeed, homeodomain proteins function *in vitro* and *in vivo* as sequence-specific transcription factors. Genetic analyses with yeast and DNA-binding studies *in vitro* have shown that, for several major classes of *Drosophila* homeodomain protein, a single amino-acid residue at position 9 of helix III — the recognition helix — is used to distinguish between related homeodomain-binding sites. Switching the identity of this amino acid switches the target specificity from one particular consensus DNA sequence to another. These results, and the suggestion that interaction between a homeodomain protein and its target sequence is direct, have now been confirmed *in vivo* in *Drosophila* for the *fushi tarazu*¹ (*ftz*) and the *bicoid* homeodomain proteins (A. Schier, Biozentrum der Universität Basel; S. Hanes, Massachusetts General Hospital, Boston).

X-ray crystallographic studies of homeodomain-DNA co-crystals², although in general agreement with these observations, differ in significant details; surprisingly, residue 9 seems to interact, through a van der Waals bond, with base pair 8 of a consensus DNA binding site, rather than with base pair 7 as predicted in the genetic model³ (see figure). On the other hand, a high-resolution NMR structure of the complex (K. Wüthrich, Institut für Molekularbiologie und Biophysik, Zürich) is consistent with the *in vivo* data but now adds the twist that two water molecules appear to be nestling between the recognition helix and the DNA, one of them directly between residue 9 and base pair 7. Although no water molecules were observed in the X-ray crystal structure, there is enough space (about 4 Å) between the bases at the bottom of the major groove and the

side-chains of helix III to accommodate them. One can imagine how specific, even discriminatory, interactions might be mediated by water molecules, given the strong angle or orientation dependence of hydrogen bonds. There is at least one precedent for this type of transmission of specific information between proteins and nucleic acids⁴, in the structure of a transfer RNA synthase complexed with tRNA.

This is not to say that residue 9 is the



Determination of the specificity of homeodomain-DNA interactions. Top: A prediction of genetic studies is that amino-acid residue 9 of helix III interacts with base pair 7 of a consensus binding site. Bottom: On the evidence of a crystal structure of the *engrailed* homeodomain, however, it seems that residue 9 (shown as Q) might interact with base pair 8 of an equivalent DNA-binding site.

only determinant of binding specificity between homeodomain and DNA; specific residues in the helix-turn-helix that contact the DNA backbone are also required for efficient recognition of the target site by *ftz* (K. Furukubo-Tokunaga, University of Basel).

Homeodomain proteins are classified, according to the relatedness of the sequence of their homeodomains, and homeodomain proteins within one class will tend to bind similar, if not identical, DNA sequences. The Antennapedia and

Ultrabithorax proteins of *Drosophila* are a case in point. Yet these proteins have markedly different functions in *Drosophila*. So it seems that specificity of function cannot wholly reside in selection of DNA-binding sites (although this is yet to be formally excluded), and that there must be other features of the proteins that affect their function. Swapping segments of these homeodomain proteins with each other, and then assaying the function of the resultant chimaeric proteins in *Drosophila*, has proved a useful technique for finding out which amino-acid residues are central for function *in vivo*. At the meeting R. Mann (Columbia University), W. McGinnis (Yale University) and M. Scott (Stanford University) described results showing that amino acids at the extreme amino terminus of the homeodomain, and those immediately adjacent to the amino and carboxy termini, are important for specificity of function. There is no indication of what precisely these residues do as yet; whether they affect DNA binding, or transcriptional activity by interacting with other factors, remain open questions.

It is here that observations on the function of two homeodomain proteins in *Saccharomyces cerevisiae* ($\alpha 1$ and $\alpha 2$) may well be of broad significance, in that they should provide clues to the mode of action of homeodomains in metazoans. Sequences immediately carboxy-terminal to the $\alpha 2$ homeodomain mediate an interaction with the $\alpha 1$ protein, and amino-terminal sequences interact with another partner in transcriptional control, MCM1 (S. Johnson, University of California San Francisco; ref. 5). Indeed the amino-terminal $\alpha 2$ sequences interact with $\alpha 1$ through the homeodomain of $\alpha 1$, lending support to the notion that the biological specificity of the homeodomain proteins may be mediated by associated proteins, and that these may themselves be homeodomain proteins.

At a different level, investigation of the influence of homeodomain proteins in patterning various tissues, in organisms as diverse as *Caenorhabditis elegans* and chickens, is gathering momentum. Alteration of the expression of some vertebrate Hox genes (homeobox genes that are found in highly conserved clusters in the genome) cause homeotic transformations of the body plan in a similar fashion to those observed in *Drosophila*. P. Brûlet (Institut Pasteur) has now added to this small but growing canon; null mutations in the mouse *Hox-3.1* gene result in the transformation of several skeletal elements into the likeness of more anterior ones⁶, as observed in *Drosophila* with loss-of-function homeotic mutations. C. Tabin (Harvard Medical School) presented evidence that

*EMBO Workshop: The Homeobox and the Genetic Control of Development, Ascona, Switzerland, 24–30 May 1992.

localized ectopic expression of a Hox gene, *Hox-4.6*, in the limb bud of the developing chick results in a homeotic transformation of one digit into another⁷. These results are the first direct demonstration that Hox genes actually control the patterning of the vertebrate limb.

How widespread is the homeobox? A screen for homeobox sequences, employing the polymerase chain reaction, reveals that Hox genes are present in the principal animal phyla and points to their being universal in animal metazoans. Moreover it seems that the different classes of homeodomain protein were established early during evolution, around the time of the rapid metazoan radiation some 600 million years ago (C. Kappen and F. Ruddle, Yale University). All these results suggest that the homeobox was a central player in the evolution of organismal form. Certainly

they are providing hints as to the answer to a long-standing puzzle for students of the vertebrate head — did the head arise by modification of ancestral segments in a headless ancestor, or was it made from scratch? The expression patterns of four homeodomain proteins in the mid- and forebrain of mice (A. Simone, CNR, Milan), the *Drosophila* homologues of which are required for development of the fly head, lend support to the first explanation. □

Guy Riddihough is an assistant editor of Nature.

1. Schier, A. F. & Gehring, W. J. *Nature* **356**, 804–807 (1992).
2. Kissinger, C. R. *et al. Cell* **63**, 579–590 (1990).
3. Hanes, S. D. & Brent, R. *Science* **251**, 426–430 (1991).
4. Rould, M. A., Perona, J. J. & Steitz, T. A. *Nature* **352**, 213–218 (1991).
5. Smith, D. & Johnston, A. D. *Cell* **68**, 133–142 (1992).
6. Le Mouelluc, H. *et al. Cell* **69**, 251–264 (1992).
7. Morgan, B. A. *et al. Nature* (in the press).

ECOLOGY

Pastoral plumbing

Peter D. Moore

SPATIAL and temporal heterogeneity in an ecosystem promotes the development of biological diversity and is therefore often sought after by those concerned with management for conservation purposes. In grasslands, for example, large grazing animals contribute to microhabitat diversification not only as a result of their trampling and biting of herbage, but also by defaecating and urinating in patches.

V. J. Jaramillo and J. K. Detling^{1,2} have examined the effects of patch urination on grasslands, both in terms of its effect on nutrient uptake and growth of grasses and also its subsequent influence on the grazing behaviour of herbivores. They find that the grasses tested were stimulated by treatment with artificial urine and that subsequent grazing by cattle was intensified within the patches. Evidently this will favour the development of structural heterogeneity in grassland, but its ultimate influence on species diversity is more difficult to predict.

Excretion by grazing animals is known to affect the internal nutrient cycling pattern of a range of different ecosystems, both terrestrial and aquatic, and nutrient redistribution and local concentration produced in this way can modify the growth and productivity of primary producers. Fish schools grazing in sea grass during the day and roosting amongst corals, for example, increase the local ammonium ion concentrations by over four times and this can stimulate coral growth because of the provision of

extra nutrients to the photosynthetic symbionts of the corals³. On an even smaller scale, copepod excretion has been shown by radioactive-labelling studies to create a spatial heterogeneity of phosphorus concentration that can result in patches of enriched growth amongst phytoplankton⁴. Similarly, on land, invertebrate grazers such as snails in the Negev desert consume endolithic lichens from rocks during the cool of the night, then deposit much of the nitrogen in the shade of those rocks during the day, creating pockets of eutrophicated soil⁵.

Jaramillo and Detling¹ looked at the development of patchy environments in grassland near Cheyenne, Wyoming, by applying treatments of two litres of simulated urine to selected plots while retaining untreated matched plots as controls. They then concentrated their attentions on two grass species, *Agropyron smithii*, a rhizomatous C3 photosynthesizer, and *Bouteloua gracilis*, a tussocky C4 species. Both species increased their tiller density as a result of treatment, *Bouteloua* in the first year and *Agropyron* in the second. The temporal difference in response could be related to the greater efficiency of nitrogen use in the C4 species; less nitrogen is needed for photosynthetic enzyme production, so more can be devoted to vegetative growth. The influence of the local fertilization by urine was felt outside the immediate ring of application — in the case of the rhizomatous species, this is likely to be a consequence of transport along the rhizomes, but movement in the

clump-forming species is more difficult to explain. Perhaps nitrogen transfer occurred by mycorrhizal links between individual plants of the same (or even different) species.

In a second experiment², Jaramillo and Detling observed the behaviour of cattle when presented with grassland in which simulated patch urination had taken place. Selective grazing occurred, in which the fertilized plots were favoured, and this corresponded with elevated foliar nitrogen concentration of the grass. This response by the grazing animals supports the ideas of Bazely, Ewins and McCleery⁶, which stem from their observations of the behaviour of barnacle geese (*Branta leucopsis*) on salt marshes in the Netherlands. The geese concentrated their grazing in areas that had been used for colonial breeding by herring gulls (*Larus argentatus*) and in which the foliar nitrogen content of the red fescue grass (*Festuca rubra*) was increased. A cycle of urination, fertilization, growth stimulation and increased grazing appeal seems now to be well established by these various studies.

The question of species diversity is more difficult, however. If we accept the Grime model of moderate stress leading to enriched diversity⁷, then local fertilization is more likely to stimulate the growth of 'competitor' plants and suppress those less robust in stature and aggressive in growth. But if those same patches are then subjected to increased grazing, the effect could be reversed, with fast-growing species being selectively removed by the herbivores. In practice, the establishment of species-rich grassland, according to the work of Gibson and Brown⁸, demands more than a quick fix by a selected grazing regime. They claim that centuries of very heavy grazing is the only sure way of establishing the species-rich, 'ancient' grasslands which are so highly prized by conservationists in temperate regions. There is no magic short-cut. Within our life spans we may have to make do with the patchy environment created by a leaky cow. □

Peter D. Moore is in the Division of Biosphere Sciences, King's College London, Campden Hill Road, London W8 7AH, UK.

1. Jaramillo, V. J. & Detling, J. K. *J. appl. Ecol.* **29**, 1–8 (1992).
2. Jaramillo, V. J. & Detling, J. K. *J. appl. Ecol.* **29**, 9–13 (1992).
3. Meyer, J. L., Schultz, E. T. & Helfman, G. E. *Science* **220**, 1047–1049 (1983).
4. Lehman, J. T. & Scavia, D. *Science* **216**, 729–730 (1982).
5. Jones, C. G. & Shachak, M. *Nature* **346**, 839–841 (1990).
6. Bazely, D. R., Ewins, P. J. & McCleery, R. H. *Ibis* **133**, 111–114 (1991).
7. Grime, J. P. *Plant Strategies and Vegetation Processes* (Wiley, Chichester, 1979).
8. Gibson, C. W. D. & Brown, V. K. *J. appl. Ecol.* **29**, 120–131 (1992).

Metallic solid silicon

Robert W. Cahn

"SILICON is the most intensely studied material in the scientific literature." While this recent assertion¹ is undoubtedly true, remarkable new aspects of silicon are nevertheless still being discovered. One such is a transition from the normal semiconducting (non-metallic) state to a metallic state in the neighbourhood of tiny plastic indentations on pure silicon crystals, created by a diamond pyramid under load²⁻⁶. This discovery, the final outcome of a long series of theoretical insights in electron theory and experiments initially in high-pressure laboratories, has about it the unmistakable scent of a major advance.

The theoretical insights referred to go back to Mott's treatment⁷ of the metal-nonmetal (MNM) transition in group IV semiconductors. This transition, from a semiconductor to a metallic state, depends on doping the semiconductor (silicon or germanium) with enough pentavalent solute for the localized atomic wavefunctions of the solute atoms to begin to overlap, screening the

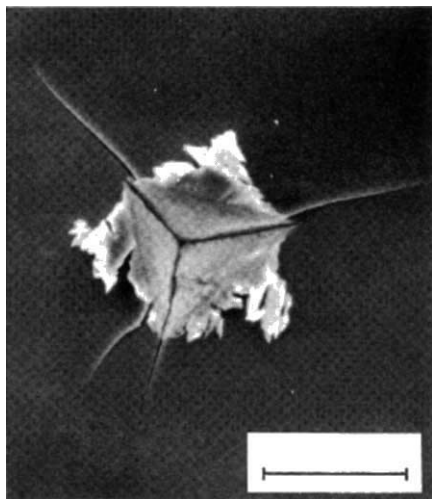


FIG. 1 Scanning electron micrograph of a 40-millinewton indentation in silicon; scale bar is 2 μm . (Courtesy of G. M. Pharr.)

extra electrons from the extra nuclear charge of the solute atoms. Mott's celebrated quantitative criterion for the MNM transition, as well as several later variants of the theory, are particularly clearly explained by Cottrell⁸, and the remarkable consistency of the theory with a wide range of experimental findings has been recorded by Edwards and Sienko⁹.

The original theory referred only to MNM transitions generated by doping, but a later version, the 'polarization catastrophe' picture^{9,10} based on pre-quantum ideas due to Goldhammer and Herzfeld, shows that an MNM transition

can equally be forced by compressing a semiconductor, increasing the atomic density until the dielectric constant diverges (a 'cooperative polarization') at a critical density.

It is well established that ordinary diamond-cubic silicon transforms to the much denser β -tin structure under hydrostatic pressure at ambient temperature; the critical pressure is in the range 11.3–12.5 GPa. This is close to the indentation hardness of diamond-cubic silicon, which is in the range 11–12 GPa. Several physicists, including Gerk and Tabor¹¹, were struck by this similarity and suggested that silicon might transform to the β -tin structure under a diamond indenter, where the stress state is a combination of hydrostatic and deviatoric components. This suggestion was consistent with an experiment¹² that had revealed a large, reversible drop in resistivity in a thin layer surrounding an indentation in silicon, as would be expected in view of the metallic character of the β -tin structure. The recognition of the similarity of the critical pressure to the indentation hardness and this experiment¹² underlay some of the recent rash of new studies . . . but serendipity played its customary part as well.

The first study, mostly carried out at IBM's Yorktown Heights laboratory, was by Clarke *et al.*². They observed the formation of an amorphous phase under a diamond pyramid pressed into a silicon crystal. This research was part of a programme to study crack formation in brittle solids. Once the unexpected phase transformation had been observed, the experimenters recalled Gerk and Tabor's suggestion and measured the electrical resistivity, finding a reversible reduction near the indentation.

Clarke next joined forces with Pharr, in Texas, and Oliver, in Tennessee³. Oliver had invented an instrument, the Nanoindenter, which allowed accurate hardness measurement to be made under exceedingly small loads. Their study of nanoindentation hardness of many materials included silicon crystals of various orientations. Two features unique to silicon were found: with (relatively) large peak loads, a sharp discontinuity in displacement was found during the unloading cycle; with smaller peak loads, there was a reversible hysteresis cycle. The discontinuity was attributed to cracking (only now is this confirmed, in Pharr's latest contribution⁶); the hysteresis, which was accompanied by a large change in resistivity, was associated with a pressure-induced phase transition.

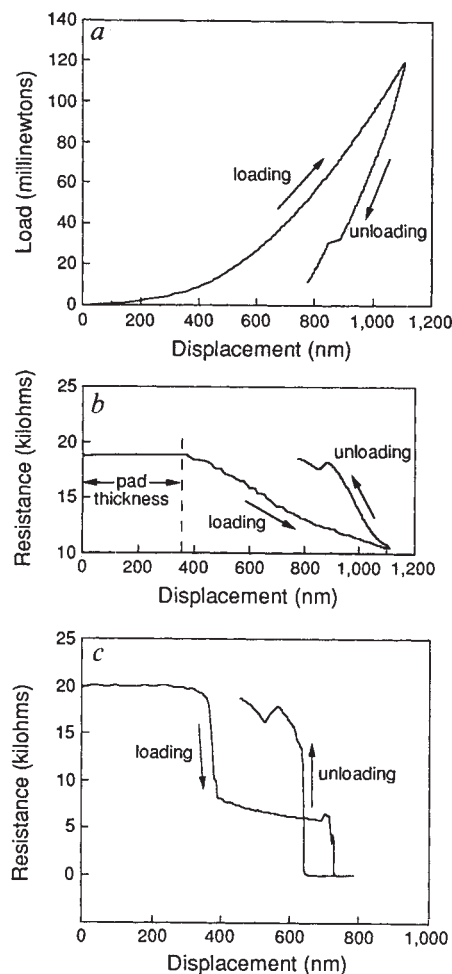


FIG. 2 Indentation of silicon. *a*, Displacement hysteresis during a loading-unloading cycle, showing the strain discontinuity during unloading. *b*, Resistivity change for an indentation through a gold electrode; the change is small because much of the current path is through unmodified silicon between adjacent electrodes. *c*, Resistivity changes when the indenter bridges a gap between electrodes, so the amorphized silicon touches both.

The next step was the observation⁴ of unmistakable plastic extrusions around nanoindentations in silicon (Fig. 1). This, again, was unique to silicon among the many brittle materials examined. The implication is that silicon transforms to a metastable form under the diamond and that the plastic behaviour, as well as the electrical properties, become those characteristic of a metal.

This preliminary study was followed by a more detailed one, in which the mechanical and electrical hysteresis were thoroughly analysed, by depositing an array of strip electrodes on the surface and indenting on or between the strips⁵. The effects of the MNM transition were seen only if an indentation pierced or touched an electrode (Fig. 2). The authors emphasize that the resistivity changes are due partly to the phase transition and partly to changes in the

characteristics of the electrode/silicon interface as the transition reaches each electrode: the rectifying Schottky barrier is replaced by an ohmic metal-to-metal interface.

Independently and very recently, Gilman reviewed¹³ MNM transitions in a range of semiconductors and insulators and showed some striking correlations between the experimental 'metallization pressures' and values calculated from the polarization catastrophe model. He predicts that many such transitions should be observable under indenters, and speculates whether plastic shear may play a role in the MNM transition. He also suggests that, alternatively, 'metallization' and dislocation motion may both be correlated with the excitation of bonding electrons into antibonding states, and offers some supporting evidence from the literature.

In the meantime, Pirouz *et al.*¹⁴ have been conducting a separate series of experiments to study the partial transformation of diamond-cubic silicon to a diamond-hexagonal form by indentation at 400–650 °C (Clarke, Pharr and colleagues' experiments were at room temperature). They show that this process is martensitic (shear-induced); here the deviatoric stress component under an indenter is crucial. But the authors have apparently not looked for an MNM transition in the hexagonal silicon.

There remains the question of how the amorphous, conducting phase² arises. Conventional pressurization experiments have hitherto shown no sign of it. This is addressed in Pharr's latest contribution⁶. He marshals the extensive published evidence to show that silica, ice and several semiconductors other than silicon become amorphous under pressure. Indeed, non-hydrostatic stresses, at the edge of an indentation, help give silica two distinct pressure-induced amorphous phases. And it is also clear that thin amorphous silicon films can undergo an

abrupt MNM transition under hydrostatic pressure¹⁵ (just as unpressurized amorphous silicon 'melts' to produce a denser, metallic melt). Perhaps silicon will amorphize only under a combination of hydrostatic and deviatoric stresses. Pharr simply concludes that "at some point in the transformation, an amor-

phous phase is formed". As people are wont to say in grant proposals, more research is clearly needed. □

Robert W. Cahn is in the Department of Materials Science and Metallurgy, University of Cambridge, Pembroke Street, Cambridge CB2 3QZ, UK.

EARTH SCIENCES

Water in mantle minerals

David R. Bell

NEW experiments suggest that olivine, probably the most abundant mineral in the Earth to a depth of 400 km, can incorporate substantial amounts of hydrogen under water-rich, high-pressure conditions. This work, reported by Bai and Kohlstedt on page 672 of this issue¹, addresses two basic questions facing those looking at the role of water in the evolution of the Earth: how much is there and where is it stored?

Geologists recognize that dynamic physical and chemical processes in the Earth's interior — such as those which dictate the distribution of the continents, control the generation and eruption of magmas, and perhaps lead to the accumulation of the atmosphere and oceans themselves — can be strongly influenced by the presence of water. Since the suggestion² that a protracted degassing of the Earth's interior gave rise to the atmosphere and oceans, the search has been on for suitable host phases for mantle hydrogen. Although current opinion favours early, perhaps catastrophic release of volatiles from the Earth's interior, the identity of H storage sites remains important for understanding water recycling and for predicting the behaviour of H during processes deep in the Earth. This knowledge should give a better appreciation of how the evolution of the atmosphere and oceans is linked to processes in the Earth's interior.

The presence of hydrous minerals such as amphiboles and micas in mantle-derived xenoliths (accidental rock fragments transported to the surface in volcanic eruptions) provides clear evidence of water storage in the upper mantle. Unfortunately, it is difficult to generalize these local observations, because the mantle is known to be heterogeneous, especially under the continents whence many of the xenoliths are derived. Many of the observed hydrous minerals may be local phenomena generated by relatively small-scale magmatic events, thereby not representing the mantle at large. As we don't have direct access to most of the mantle, our understanding must rely heavily on laboratory studies.

Experimental phase-equilibrium studies which simulate conditions in the Earth's interior can place powerful constraints on the expected phase assemblages in the upper mantle. Studies on compositions in the mantle-analogue system, MgO–SiO₂–H₂O, have produced a number of hydrous phases which are candidates for storing water in the mantle. However, many of these phases are apparently not stable at the high temperatures of average mantle and thus would be restricted to anomalously cold regions such as subducting slabs.

Minor components not accounted for in simple synthetic systems may stabilize small quantities of hydrous phases, which, because of the generally water-poor nature of the mantle (probably 0.01–0.1 per cent H₂O by weight), may be relatively important water reservoirs. Experimental studies on micas and amphiboles have revealed that certain compositional variants of these phases (particularly those rich in potassium and fluorine) are stable well into upper-mantle pressure and temperature conditions. Similarly, titanium seems to stabilize several of the humite group minerals (structures composed of alternating layers of olivine and brucite [Mg(OH)₂]) to mantle pressures and temperatures³.

In general, the incompatible-element content of the upper mantle is low, and large concentrations of K, Ti or F-rich hydrous minerals cannot be supported. As a result, such hydrous minerals may be restricted to areas locally enriched in these elements, with the nature of the water storage sites depending on the scale of mantle chemical heterogeneities.

Here at Caltech, Rossman and co-workers have found small, but virtually ubiquitous quantities of H in the common, nominally anhydrous mantle minerals, such as olivine, garnet and the pyroxenes (see ref. 4). Hydrogen is bound in the mineral structure as hydroxyl groups (OH). Concentrations vary from less than 1 part per million (p.p.m.) up to 0.1 per cent H₂O by weight. The effect on the transport properties of the host phase⁵, with its attendant implications for mantle rheology,

- Huff, N. R. in *Concise Encyclopedia of Semiconducting Materials and Related Technologies* (eds Mahajan, S. & Kimerling, L.) 478–492 (Pergamon, Oxford, 1992).
- Clarke, D. R., Kroll, M. C., Kirchner, P., Cook, R. F. & Hockey, B. J. *Phys. Rev. Lett.* **60**, 2156–2159 (1988).
- Pharr, G. M., Oliver, W. C. & Clarke, D. R. *J. electron. Mater.* **19**, 881–887 (1990).
- Pharr, G. M., Oliver, W. C. & Harding, D. S. *J. Mater. Res.* **6**, 1129–1130 (1991).
- Pharr, G. M. *et al.* *J. Mater. Res.* **7**, 961–972 (1992).
- Pharr, G. M. in *Thin Films: Stresses and Mechanical Properties III* (eds Nix, W. D. *et al.*) (Materials Research Society, Pittsburgh, in the press).
- Mott, N. F. *Proc. Phys. Soc.* **A62**, 416 (1949).
- Cottrell, A. *An Introduction to the Modern Theory of Metals*, 16–20 (Institute of Metals, London, 1988).
- Edwards, P. P. & Sienko, M. J. *Phys. Rev.* **B17**, 2575–2581 (1978).
- Edwards, P. P. & Sienko, M. J. *Int. Rev. phys. Chem.* **3**, 83 (1983).
- Ger, A. P. & Tabor, D. *Nature* **271**, 732 (1978).
- Grideva, I. V., Millman, Y. V. & Triflov, V. I. *Phys. Stat. Sol.* **14**, 177 (1972).
- Gilman, J. J. *J. Mater. Res.* **7**, 535–538 (1992).
- Pirouz, P., Chaim, R., Dahmen, U. & Westmacott, K. H. *Acta metall. Mater.* **38**, 313–322; 323–328; 329–336 (1990).
- Shimomura, O. *et al.* *Phil. Mag.* **29**, 547 (1974).

MALARIA

Asexual deviants take over

Russell J. Howard

have motivated a number of studies of the olivine-H₂O system, including the present one by Bai and Kohlstedt. The nominally anhydrous minerals are the dominant constituents of the mantle, so they are potentially important reservoirs of H in the Earth's interior.

Hydrogen concentrations measured in minerals from natural mantle samples suggest OH-bearing capacities sufficient to account for the water content^{6,7} of the depleted upper mantle (about 100–200 p.p.m.). Although the measurements on natural samples are valuable in demonstrating the ability of these phases to store OH, the concentrations are not easily generalized to average mantle. Because of the low H concentrations, relatively little is known about the substitution mechanisms by which OH is incorporated in these phases, where it sits in the mineral structure or how the concentrations are affected by pressure, temperature and a host of extensive parameters encountered in the mantle.

Bai and Kohlstedt have taken the first step in erecting a quantitative experimental framework for the understanding of the behaviour of H in this reservoir. Their study demonstrates that the hydrous component in olivine is stable to a temperature of 1,300 °C at relatively low pressure, thereby exhibiting considerably greater thermal stability than most hydrous magnesium silicates. If the hydrous components of other nominally anhydrous minerals, in particular pyroxenes, which appear⁴ to contain the most OH of the common mantle minerals, behave similarly then these minerals may be able to store H under conditions where most stoichiometrically hydrous minerals break down.

It will be important to investigate the effect of pressure on the solubility of H in olivine and other mantle minerals, as Bai and Kohlstedt's experiments were conducted at pressures of the uppermost crust, and require considerable extrapolation to the mantle regime. As OH concentrations in olivine appear to be linked with certain point defects, it remains to be seen if saturation effects or other substitution mechanisms come into play at high water fugacities and perturb the reported systematics. □

David R. Bell is in the Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, California 91125, USA.

THE blood stage of malaria is a promising target for the development of new antimalarial drugs and vaccines. At this stage the parasite enters erythrocytes where it multiplies asexually. Infected erythrocytes express various cell-surface molecules which can provoke immune responses and which could also be targets for antimalarial drugs. But work reported in this issue by Roberts *et al.* (page 689)¹ illustrates a particularly intriguing case of the way in which parasites evade the body's defences by variation in these cell-surface molecules and so make the task of control more difficult.

The life of an asexual bloodstage malaria parasite is a continuous battle for survival. *Plasmodium falciparum*, the most deadly of the four human malarias, exploits two strategies to survive both innate and acquired immune responses. Mature malaria-infected erythrocytes express receptors for endothelial cells that allow them to adhere in microvessels and prevent them from passing through the spleen, where they are likely to be destroyed by the filtration mechanism in that organ. Infected erythrocytes sometimes express receptors for uninfected red blood cells; rosettes are formed, and these are also sequestered from the peripheral circulation and protected from destruction in the spleen. Malarial proteins expressed on infected erythrocytes have been identified as candidate adhesion receptors. Altered erythrocyte membrane proteins may also participate in adherence. Although the appropriate malarial genes have not been cloned, it is becoming increasingly clear that there is not just a single receptor specificity shared by all parasite isolates.

The second mechanism of immune evasion arises from the antigenic diversity of the parasite population, with each isolate expressing different cell-surface antigen(s) on infected erythrocytes. Antigenic variation in the progeny of a single parasite has also been observed *in vitro*². The combined effects of antigenic diversity in the population and antigenic variation in clonal progeny make it likely

that there will always be some parasites able to circumvent pre-existing immunity in the host.

Research on these two mechanisms for evading host immunity has now converged. Roberts *et al.*¹ show that, in the absence of immune pressure, *P. falciparum* can vary the antigenic phenotype of surface antigen and the specificity of adhesion receptor(s) on infected cells at an alarming frequency of some 2% per generation. It also seems that particular serotypes of the variant antigen correlate with particular combinations of receptor



Sticky business: a COS-7 cell, transiently transfected with CD36, binds red blood cells infected with *P. falciparum*.

specificities.

Several purified human proteins known to be expressed on the surface of endothelial cells can mediate adherence of infected erythrocytes; these include thrombospondin (ref. 3), CD36 (ref. 4), ICAM-1 (ref. 5), VCAM-1 and ELAM-1 (ref. 6). Different strains or clones of cells infected with *P. falciparum* express receptors for only some of the endothelial cell adhesion molecules; thus, infected erythrocytes that adhere to CD36, for example, may or may not also adhere to ICAM-1 (ref. 5). The genetic origin of receptor diversity is unknown. Does it represent spontaneously derived de-

1. Bai, Q. & Kohlstedt, D. L. *Nature* **357**, 672–674 (1992).
2. Rubey, W. W. *Geol. Soc. Am. Bull.* **62**, 1111 (1951).
3. Khodyrev, O. Y., Agoshkov, V. M. & Slutskiy, A. B. *Trans. USSR Acad. Sci.* **312**, 255–258 (1990).
4. Bell, D. R. & Rossman, G. R. *Science* **255**, 1391–1397 (1992).
5. Mackwell, S. J., Kohlstedt, D. L. & Paterson, M. S. *J. geophys. Res.* **90**, 11319–11333 (1985).
6. Michael, P. *Geochim. Cosmochim. Acta* **52**, 555 (1988).
7. Dixon, J. E., Stolper, E. M. & Delaney, J. R. *Earth planet. Sci. Lett.* **90**, 87–104 (1988).

variants continuously generated from the parent, or does it represent the existence of genetically independent alternative phenotypic forms present as minor contaminants in non-clonal parasite populations?

To study this problem Roberts *et al.*¹ prepared 27 clones of *P. falciparum* blood-stage parasites from a cloned culture-adapted parent and compared them for infected erythrocyte surface antigen phenotype as well as for adherence to CD36 or ICAM-1, rosetting with uninfected erythrocytes and autoagglutination. Ten new surface-antigen phenotypes were described in cloned parasites derived from the parent clone. The authors also found modulation of endothelial receptor phenotypes in the cloned parasites — a phenomenon not discovered before. There were numerous examples of modulation such as rosetting-negative to rosetting-positive, autoagglutination-negative to autoagglutination-positive, and ICAM-1 and CD36 adherence-negative to ICAM-1 and CD36 adherence-positive.

One of the most fascinating observations in this report is that clones with the same surface-antigen phenotype have the same combination of receptor phenotypes. Does this imply that the malaria parasite uses an antigenically variant structure on the infected erythrocyte surface to fulfil a functional role as receptor for endothelial cells and erythrocyte rosettes? I have speculated earlier that the opposite must be true — that is, that the receptor for endothelial cells would be antigenically conserved, in order to fulfil function, whereas the variant surface antigen would function only in immune evasion⁷. With the results of Roberts *et al.*, it is now likely that the variant surface antigen is in fact the receptor for endothelial cells. They have demonstrated an extremely rapid rate of spontaneous generation of receptor and antigenic variants, have evidence for a correlation between antigen and receptor phenotypes, and show that numerous endothelial cell proteins can be used *in vitro* to select infected erythrocytes with a corresponding receptor specificity.

How do the asexual deviant parasites take over and dominate the parasite population in immunized hosts? The initial inoculum of asexual parasites into the blood (derived ultimately from sporozoites inoculated by the mosquito and having passed through intermediate stages in the liver) will include different phenotypes of the surface antigen and the receptor(s) for endothelial cells. Different phenotypes will be present even if only one sporozoite was originally inoculated owing to the generation of 2% deviant organisms per generation. Two selective forces will operate at the level of the surface phenotype of in-

fectured erythrocytes. Variant antigen phenotypes that had previously infected the individual will elicit specific IgG antibodies that lead to destruction of those infected erythrocytes. Antigen phenotypes never seen by the host will survive by negative selection. If the variant antigen and the receptor phenotype are closely linked genetically (or perhaps are even the same structure), the remaining parasites will express a diversity of receptor phenotypes that will then be acted on by the positive selective force for adherence to microvascular endothelial cells and/or rosette formation. Only those infected cells with receptors for tissue proteins expressed at sufficient level in the host will survive. The endothelial cell-surface proteins used for adherence are expressed differentially depending on diverse autocrine and endocrine effects including cytokine stimulation. So the proteins available for positive selection for adhesion may even vary during the course of a single infection, as the host-parasite battle elicits different cytokines and other effectors of endothelial cell-surface phenotype. Given the capacity of infected erythrocytes to produce antigenic variants and receptor deviants, and dynamic changes in the negative selective forces (antibodies) and positive selective forces (endothelial cell-surface proteins) during the course of infection, the infected erythrocytes that proliferate in each person probably have quite different surface phenotypes at different times during the disease.

Pronounced sequestration of infected erythrocytes in cerebral microvessels occurs in only some malaria cases and causes the special neurological pathology of human cerebral malaria⁸. Whether this pathology ensues will depend not only on the potential of the parasite inoculum to produce deviant parasites that adhere preferentially in cerebral bloodvessels, but on the particular antibody responses and the particular collection of endothelial cell proteins available in the individual. □

Russell J. Howard is at DNAX Research Institute of Molecular and Cellular Biology, Palo Alto, California 94304-1104, USA.

1. Roberts, D. J. *et al.* *Nature* **357**, 689–692 (1992).
2. Biggs, B. A. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 9171–9174 (1992).
3. Roberts, D. D. *et al.* *Nature* **318**, 64–66 (1985).
4. Barnwell, J. W. *et al.* *J. Clin. Invest.* **84**, 765–772 (1989).
5. Berendt, A. R., Simmons, D. L., Tansey, J., Newbold, C. I. & Marsh, K. *Nature* **341**, 57–59 (1989).
6. Ockenhouse, C. F. In *Proc. Second Wellcome Foundation Meeting on Severe Malaria*, Broadway, Worcestershire, 16–19 Feb., 1992.
7. Howard, R. J. *Prog. Allergy* **41**, 98–147 (1988).
8. MacPherson, G. G., Warrell, M. J., White, N. J., Looareesuran, S. & Warrell, D. A. *Am. J. Pathol.* **119**, 385–401 (1985).

DAEDALUS

Hot stripping

LATHES and milling machines face a subtle contradiction. The workpiece must be rigid enough not to deflect under the cutting tool, yet the cuttings must be flexible enough to bend out of its way. The solution is to take repeated fine cuts from a relatively massive workpiece. But very soft and very hard materials defeat the machinist. Rubber deforms hopelessly away from the cutter, while ceramics refuse to deform even as fine cuttings; they simply shatter.

This contradiction between the rigid workpiece and the flexible cuttings can sometimes be evaded. Soft rubber can be cleanly drilled or machined by first cooling it in liquid nitrogen. It becomes as rigid and brittle as glass, yet does not shatter under the drill or cutter. The reason is simple. Whereas the bulk of the rubber is unmechanically brittle, its surface is warmed into flexibility by the air and the heat of machining. Supported on the rigid interior, it curls off cleanly and smoothly, without tearing or sagging away from the cutter.

So Daedalus is repeating this trick at a higher temperature. Imagine, he says, a standard centre-lathe with a powerful heater mounted just ahead of the cutting tool, and playing on the fast-spinning workpiece beneath. The heater will 'write' a continuous heated strip onto the moving surface. If sufficiently intense and localized, this hot strip will not cool and spread by conduction in the millisecond or so before it is 'read' by the cutter a few millimetres downstream. The bulk of the workpiece will remain cool and rigid. But the cutter will be continuously presented with a softened, red-hot skin, easily detached and peeled away from the firm interior.

Few heaters are intense and local enough for this duty. For metallic, electrically conducting workpieces, an arc-welding electrode seems feasible. Daedalus likes the idea of recycling the heat of cutting, by guiding the red-hot strip of freshly cut metal back onto the metal about to be cut; but the heat transfer would probably be too poor.

His real goal, however, is to machine hard, brittle materials such as glass and ceramics. Such insulating and non-combustible substances would have to be spot-heated by a powerful focused laser. Then, like liquid-nitrogen-cooled rubber, they would yield to the lathe. Cups, saucers, plates and holloware could be machined to close tolerances; more significantly, so could big telescope mirrors. And a numerically controlled laser-heated milling machine could form anything from rococo cut-glass decanters to ceramic turbine blades of matchless perfection.

David Jones

Thermostat and global warming

SIR — Deep convection over the tropical oceans is triggered primarily when sea surface temperatures (SSTs) exceed about 300 K. The central question is why the maximum SSTs are within a few degrees of the convection threshold temperatures in regions of convection such as the western Pacific warm pool. According to our thermostat hypothesis¹, deep convection gives rise to thick anvil clouds reflecting solar radiation back to space and limiting maximum SSTs. Wallace proposes a competing mechanism² in which such a thermostat is not required. In our view it is difficult to fit the available observations in the context of this mechanism.

Wallace explains his ideas in terms of the efficiency of tropical dynamics, which maintain spatially uniform temperatures in the troposphere. This dynamical effect is also a fundamental link in our thermostat mechanism; without it the greenhouse effect of the anvil would have balanced locally a major fraction of the surface cooling due to solar reflection and its effectiveness as a thermostat would be much weaker.

Wallace asserts that even without the reflective clouds, the evaporative cooling will increase with SST to limit SSTs to the observed values. How do we test the validity of the two hypotheses? The best recourse is to test their predictions with available observations. Here, we examine observational estimates for evaporative heat flux on several time and spatial scales. Our interest is in those oceanic regions where deep convection and warm SSTs (>300 K) occur together.

Climatological estimates^{3,4} reveal that the evaporation heat flux decreases towards the warm pool (see figure). Further, its spatial gradient shown in the figure would tend to amplify SST gradients. The absorbed solar radiation gradient, on the other hand, tends to decrease the SST gradient by having a cooling effect over regions of convec-

tion, as it decreases significantly towards the warm pool.

Wallace's proposal involves 'hot patches' in the ocean giving rise to enhanced evaporation. Hot patches form in the central equatorial Pacific during El Niño events and are accompanied by deep convection. Estimates of evaporation from satellite data for winds and humidities reveal that the warming is accompanied by decreased evaporation from the central equatorial Pacific⁵. At the same time the absorbed solar energy decreases significantly¹; this decrease lags behind the SST increase by about 2 months⁶, an important delay implying a link between SST, deep convection, cloudiness and the large-scale atmospheric dynamics as suggested in our paper¹. Even on monthly timescales, limited observations⁷ in the warm pool suggest that the response of evaporation to SST changes accounts for only a small fraction of the total heat flux variances in regions of convection.

Thus, for the monthly to yearly timescales relevant to the coupled system, the available observations are at variance with the evaporation hypothesis of SST equilibration. The picture that emerges is that, in regions where the warm SSTs and deep convection occur together, evaporation decreases with SST. The implication is that SST is only one of many variables that control evaporation from the ocean surface. Atmospheric boundary layer humidity is another controlling variable. The boundary layer humidity increases significantly with SST in regions of convection, thus limiting the increase of evaporation with SST³⁻⁵. In addition, the net outgoing long-wave emission at the surface or at the top of the atmosphere decreases with SST because of the 'super greenhouse' effect¹, which would also amplify the warming. Thus, we are led to a thermostat-type mechanism involving clouds to explain the equilibrium state. Another independent three-dimensional

atmospheric model simulation shows that, in the central Pacific during El Niño events, SSTs are negatively correlated with solar insolation at the sea surface, and that changes in solar radiation rather than changes in evaporation dominate the heat flux variance at the sea surface⁸.

We are not questioning the importance of evaporative cooling to the surface energy balance. At issue is the sign and magnitude of the feedback between SST and evaporation in regions of convection such as the warm pool. How warm would the sea surface be compared with observed values without the cloud reflection of solar energy? Wallace believes it would be about the same². A recent study⁹ using a dynamical ocean model coupled to a simple atmospheric model indicates that the warm pool SST is very sensitive to heat flux anomalies. A 10 W m^{-2} anomaly produces a 1 K change in the model SST. The reduction in solar radiation due to the clouds in the warm pool is of the order of $50\text{--}100 \text{ W m}^{-2}$ (refs 1,6). The maximum SST derived by us¹ is for the present climate and we have yet to establish its validity. It is premature to extrapolate our thermostat to the global warming problem, a point which we did not make explicit in our earlier paper¹. However, the radiative-convective linkages in the thermostat hypothesis, if proved, would become vital pieces of the puzzle of global warming.

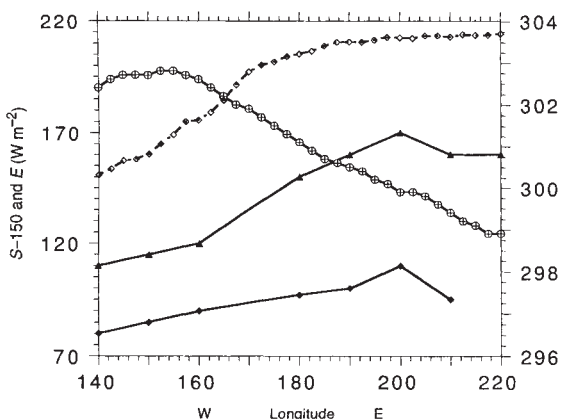
V. Ramanathan

W. Collins

Center for Clouds, Chemistry and Climate,

Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

1. Ramanathan, V. & Collins, W. *Nature* **351**, 27–32 (1991).
2. Wallace, J. M. *Nature* **357**, 230–231 (1992).
3. Weare, B. C., Ted Strub, P. & Samuel, M. D. *J. phys. Oceanogr.* **11**, 705–717 (1981).
4. Reed, R. K. *J. Clim. appl. Meteorol.* **24**, 833–840 (1985).
5. Liu, T. W. & Gautier, C. *J. geophys. Res.* **95**, 13209–13217 (1990).
6. Chertock, B. & Frouin, R. Somerville, R. C. *J. Climate* **4**, 639–650 (1991).
7. McPhaden, M. J. & Hayes, S. P. *J. geophys. Res.* **96**, 3331–3342 (1991).
8. Barnett, T. P., Kirk, E., Latif, E. & Roeckner, E. *J. Climate* **4**, 487–515 (1991).
9. Gent, P. R. *J. geophys. Res.* **96**, 3323–3330 (1991).



Annual mean values averaged between 2.5° N and 2.5° S . S , solar energy absorbed by the surface-atmosphere column, is measured by the Earth Radiation Budget Experiment (ERBE). We subtract 150 W m^{-2} from S to plot it on the same scale as evaporation heat flux, E . S (open diamonds) and SST (open circles) are observed values for 1985 and are from ref. 1. Climatological estimates of E are obtained from ref. 3 (black triangles) and ref. 4 (black diamonds).

Scientific Correspondence

Scientific Correspondence is a relatively informal section of *Nature* in which matters of general scientific interest, not necessarily those arising from papers appearing in *Nature*, are published. Because there is space to print only a small proportion of the letters received, priority is usually given according to general interest and topicality, to contributions of fewer than 500 words, and to contributions using simple language.

Network-modifying ions in glass

SIR — Vessal *et al.* claim¹ to have found evidence in a computer-simulated model for microsegregation of network-modifying ions in glasses and also for high coordinations (5–6) of non-bridging oxygens around the network-forming cations, and vice versa. However, a completely different interpretation can be placed on their simulation results.

Vessal *et al.* find that the positions of the peaks for M–Si correlations (where M is Na or Rb) correspond to bond angles subtended at the intervening oxygen atoms of θ (M–O–Si) $\sim 100^\circ$; such small angles are taken to imply that the average coordination number of such O atoms is considerably larger than that characteristic of normal bridging oxygen sites (2). Further, the average coordination of all types of O in the simulated structure is ~ 3 ; since 60% of the O atoms are bridging, Vessal *et al.* therefore conclude that the coordination of the remaining non-bridging oxygens is ~ 5 (ions), and hence, the coordination of non-bridging oxygens around the M^+ ions is also ~ 5 by charge balance.

But these data are also consistent with there being, on average, about one non-bridging oxygen in the nearest-neighbour coordination shell of the M^+ ions. Let us assume that in fact there is only one nearest-neighbour non-bridging oxygen to an M^+ ion; all the other (4–5) oxygen atoms surrounding the ion are therefore bridging, each covalently bonded to two Si atoms. If the most probable Si–O–Si bond angle is taken to be $\sim 155^\circ$ (ref. 1) and if it is assumed for simplicity that the M^+ ions lie at the bisector of the Si–O–Si bond angle (on the side away from the Si atoms) and that all four such Si, O and M atoms are coplanar, the average M–O–Si bond angle is then 102° , very close to that observed in the model. Further, if we assume that the total average oxygen coordination of the M^+ ions is 6 (five bridging oxygens and one non-bridging oxygen), because the effective coordination of each bridging oxygen is in fact 3 (2 Si and 1 M) and that of each non-bridging oxygen is 2 (1 Si and 1 M), the overall average oxygen coordination number is 2.83, again very close to that observed.

The first peak in the Na–O and Rb–O partial radial distribution functions (RDFs) in ref. 1 is markedly asymmetric, with a pronounced tail at high r . The peak maximum in both cases occurs at a distance corresponding to the sum of ionic radii for cation ($r(\text{Na}^+) = 0.97 \text{ \AA}$, $r(\text{Rb}^+) = 1.47 \text{ \AA}$) and doubly charged oxygen ($r(\text{O}^{2-}) = 1.32 \text{ \AA}$). However, it is significant that the mid-point of the high- r tails corresponds to distances involving singly-charged oxygen ions ($r(\text{O}^-) = 1.76$

$\text{\AA}$), that is $r(\text{Na–O}) = 2.73 \text{ \AA}$ and $r(\text{Rb–O}) = 3.2 \text{ \AA}$. There is evidence² that partial charge transfer occurs in $\nu\text{-SiO}_2$, and thus we can associate O^- configurations with bridging sites, and O^{2-} configurations with non-bridging sites. Thus, analysis of the shape of the first peak in the alkali-oxygen partial RDF is also consistent with the present interpretation of a mixed first coordination shell containing both bridging and non-bridging oxygens. In our molecular dynamics simulation² of lithium silicate allowing for charge transfer, such a mixed Li–O coordination shell was found.

On the matter of cation microsegregation, Vessal *et al.* claim that the observation of first (O), second (M), third (O) and fourth (M) nearest neighbours

around every alkali ion demonstrates clustering of the ions exceeding that which would be expected on the basis of an even distribution of the ions throughout the structure. The partial RDF for correlations between interstitial voids in a model of $\nu\text{-SiO}_2$ exhibits (broad) peaks at ~ 4 and $\sim 8 \text{ \AA}$, and so even in the simplest model for an alkali-modified silicate glass, in which the M^+ cations randomly occupy the interstices in the silicate framework, one would expect cation–cation correlations at distances comparable to those found experimentally and in the simulation.

S. R. Elliott

Department of Chemistry,
University of Cambridge,
Cambridge CB2 1EW, UK

1. Vessal, B. *et al.* *Nature* **356**, 504–506 (1992).

2. Alavi, A., Alvares, L. J., Elliott, S. R. & McDonald, I. R. *Phil. Mag.* **B65**, 489–500 (1992).

What is a chaperonin?

SIR — Ursic and Culbertson¹ assert that yeast and *Drosophila* TCP1 proteins lack two attributes generally shared by chaperonins (*sic*) in that they are neither members of multigene families nor heat-shock proteins. These authors have confused the terms 'chaperonin' and 'molecular chaperone'. 'Chaperonin' re-

feres. Some, but certainly not all, molecular chaperones are heat-shock proteins. The *E. coli* *groE* operon is a good counterexample to the statement that chaperonins or chaperones are typically encoded in multigene families.

The report² to which Ursic and Culbertson referred proposed that TCP1 and TF55 may function as molecular chaperones. The question of possible evolutionary relatedness of TCP1 and chaperonin-60 genes was also raised in that report and it was concluded that no sequence similarity existed (see Table 1 of ref. 2). It is clear from inspection of TCP1 and chaperonin-60 amino-acid sequences, however, that similarities exist between these two proteins. These similarities are present in register throughout both types of sequences. The figure shows one region of similarity. The chaperonin-60

sequences presented were chosen to represent a wide range of phyla. These similarities are statistically highly significant according to the criteria of Schuler *et al.*³. These observations suggest a common origin for the TCP1 and chaperonin-60 genes.

Sean M. Hemmingsen

Plant Biotechnology Institute,
National Research Council Canada,
110 Gymnasium Place, Saskatoon,
Saskatchewan, Canada S7N 0W9

1. Ursic, D. & Culbertson, M. R. *Nature* **356**, 392 (1992).

2. Trent, J. D. *et al.* *Nature* **354**, 490–493 (1991).

3. Schuler, G. D. *et al.* *Proteins* **9**, 180–190 (1991).

		* + + + + + + + + + +	
Dm TCP1	...DIGDVTITNDGATILRLLEVEHP...	AAKVLVELAQLQ	85
Cg TCP1	...DIGDVTITNDGATILRLLEVEHP...	AAKVLCELADLQ	82
Sc TCP1	...DIGDVTITNDGATILRLLEVEHP...	AGKILVELAQQQ	90
Bn Cpn60 β	...KYGSPRIVNDGVTYAREVELEDVENVIGAKLVQAQAKT		77
Bn Cpn60 α	...EPGSPKPVNDGVTIARAIELPDAMENAGALIREVASKT		75
Ec Cpn60	...SFGAPTITKDGVSAREIELEDKFPENMGAMVKEVASKA		75
Hs Cpn60	...SWGSPKVTTRDGVTVAKSIDLKDQYKNIAGKLVQDVANNT		77

		* + + + + + + + + + +	
Dm TCP1	DEEVGDGTTTSVVIIAELLKNADELVKQKIHTPTSIISGY...		124
Cg TCP1	DKEVGDGTTTSVVIIAELLKNADELVKQKIHTPTSVISGY...		121
Sc TCP1	DREIGDGTTSVVIIAELLKNADELVKQKIHTPTTIITGF...		129
Bn Cpn60 β	NDLAGDGTTSVVIIAELLKNADELVKQKIHTPTTIITGF...		116
Bn Cpn60 α	NDSAGDGTTSVVIIAELLKNADELVKQKIHTPTTIITGF...		114
Ec Cpn60	NDAGDGTTSVVIIAELLKNADELVKQKIHTPTTIITGF...		114
Hs Cpn60	NEEAGDGTTSVVIIAELLKNADELVKQKIHTPTTIITGF...		116

Aligned TCP1 and chaperonin-60 amino-acid sequences (incomplete). The numbers refer to Dm (fruitfly), Cg (hamster), Sc (budding yeast), Bn (*Brassica napus*), Ec (*E. coli*), Hs (human). GenBank accession numbers are M21159, M34665, M21160, M35600, M35599, X07850 and M34664, respectively. The alignment was made using the MACAW program³. Asterisks, positions at which all seven sequences have the identical amino acid. Crosses, positions at which at least two TCP1 sequences and two chaperonin-60 sequences have the identical amino acid.

fers to ubiquitous, sequence-related proteins, including the Rubisco subunit-binding protein of plants and the GroE proteins of *Escherichia coli*. These chaperonins are a type of 'molecular chaperone', a protein required for the post-translational folding, targeting or assembly of other proteins but which does not itself form part of the final assembled structure.

Even allowing for a confusion between these two terms, however, neither of the attributes mentioned by Ursic and Culbertson are general characteristics of either chaperonins or molecular chap-

Pharmaceutical feuds

Charles Medawar

Merck v. Glaxo: The Billion-Dollar Battle. By Matthew Lynn. Heinemann/Mandarin: 1991/1992. Pp. 244. £17.99 (Hbk)/ £5.99 (pbk).

The Aspirin Wars: Money, Medicine and 100 Years of Rampant Competition. By Charles C. Mann and Mark L. Plummer. Knopf: 1992. Pp. 420. \$25.

I HAVE only once picketed a pharmaceutical company — to protest about its roaring trade in appetite stimulants in developing countries. In the mid-1980s, the company dominated this unsavoury market with a range of products based on ciproheptadine, an antihistamine drug normally used to treat minor allergic conditions. Ciproheptadine can stimulate the appetite as a side effect, but it is not without risks; it is also expensive and makes no real contribution to health. In any case, it seemed indefensible to be selling appetite stimulants in countries where so many people were malnourished and starved.

The company said at the time that it had fully reviewed the situation and had concluded that appetite stimulation was a "well-established and medically valid" use of the drug. To others this signalled a marked lack of intellectual hygiene as well as bad taste — and it would seem natural to suspect that this said something about the culture of the company as a whole.

Matthew Lynn, however, has a great many nice things to say about Merck, and nothing at all critical. His is a journalist's description of a company that anyone could feel proud to work for; it is the biggest multinational drug company in the world and regarded by many as the best. Lynn represents the company as a first-rate institution, an honourable and excellent health-care provider that also performs outstandingly as a business. Indeed, he mentions twice that Merck Sharp & Dohme has been voted the most admired corporation in the United States, although it is not made clear by whom.

Given this eulogy, it is surprising to find that Lynn represents Glaxo as just about everything that Merck would never want to be. Glaxo, he predicts, will go the way of Hoffman La Roche, the originators of Valium and other benzodiazepine tranquillizers. Lynn characterizes Roche as a flawed enterprise, a company crippled in the end by its own "poverty of ideals, humanity and humility". This is strong stuff and gives food for thought, although the reality is probably greyer than Lynn suggests. His jaundiced view of Glaxo seems barely more convincing than the rosy appreciation of Merck Sharp & Dohme.

The author supports his essential proposition — Merck good, Glaxo not — mainly through a biography rounded off by an interview with Merck's man at the top. Roy Vagelos is characterized as "the philosopher king", "one of the most brilliant American scientists of his generation", an outstanding manager and communicator who is "passionate about his commitment to corporate social responsibility". According to Lynn, his striving for honest dealing and better human values has left no part of Merck untouched. Vagelos himself says: "I do not have to worry that our people are out there describing our products in a way that I would not be proud of, in a way that I would not be proud to see on the front page of the *New York Times*".

By contrast, Sir Paul Girolami of Glaxo is represented as "dark and brooding", a brilliant but overbearing sales and numbers man. His company, it seems, is one in which morals count for rather less and where morale can dip quite low. Glaxo is depicted as secretive and opportunistic — not nearly as clever as Merck at inventing original and useful new drugs, although very sharp when it comes to selling and pricing them.

Lynn suggests that this explains the phenomenal success of Glaxo's anti-ulcer

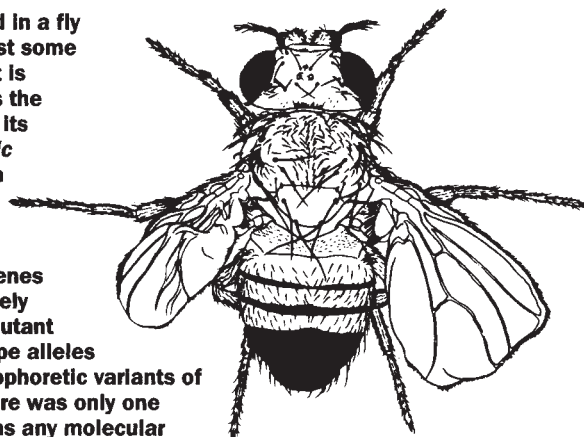
drug ranitidine: the promotion of "Zantac" turned Glaxo from a sleepy and undistinguished concern into a company whose turnover now ranks second only to that of Merck Sharp & Dohme. The gist of the story is that Glaxo used daring and cunning to establish Zantac at the top of the market, even though it was essentially an imitation of the original drug in this class, cimetidine or "Tagamet".

On therapeutic grounds, there would be little to choose between Zantac and Tagamet, yet Glaxo managed to see its rival off by persuading doctors and the public otherwise. The company took enormous financial risks by attempting to do the tests needed for licensing approval all at once, rather than in the usual strictly defined stages. In this way, Glaxo gained a lead of several years over its competitors. Then the company poured money into promotion, not least to contrast the well-established side-effect profile of the older drug, Tagamet, with the virginal record of its own product. But its masterstroke was to pitch the price of Zantac high, an audacious ploy which helped to persuade doctors that Zantac really was the better product.

Battles of the Merck versus Glaxo kind differ from those over aspirin and the like. Drugs such as Zantac compete in monopoly markets, whereas the patents on aspirin have long since expired. So the competition between manufacturers of aspirin and other over-the-counter analgesics such as ibuprofen and paracetamol/acetaminophen, has traditionally centred on the sale of image rather than substance.

Unlike Lynn's book, which is an easy read, *The Aspirin Wars* gets off to a long

Anyone who has worked in a fly lab will have had at least some acquaintance with what is affectionately known as the 'Red Book', or to give it its official title, *The Genetic Variations of Drosophila melanogaster*, by D. L. Lindsley and E. H. Grell. At the time of its publication, in 1968, genes were identified exclusively from the existence of mutant alleles; the only wild-type alleles considered were electrophoretic variants of a few enzymes; and there was only one gene for which there was any molecular information (*bobbed*). Molecular biology has now shifted the emphasis to normal gene structure and function, and this is reflected in the contents of the new edition of this fly-geneticist's Bible, *The Genome of Drosophila melanogaster*, compiled by Lindsley and G. G. Zimm. This essentially complete catalogue (up to 1989) of around 4,000 genes and 9,000 chromosome rearrangements contains some beautiful illustrations that are primarily the work of Edith M. Wallace, the artist originally employed by the Nobel laureate T. H. Morgan. Shown here is her drawing of the *humpy* mutant. Published by Academic, \$79, £51.



G.R.

and slow start. Readers must get deep into it before emerging in the twentieth century and have to wrestle with a wealth of legalistic detail about disputes over the ownership of, and the trade in, Bayer Aspirin. Brain cells expire by the billion as one tries to register and link countless names.

The early chapters are lucidly written, but seem documented far beyond the requirements of scholarship. So what if the corporate vice-president of the Bayer Company in 1917 was "Hermann C A Seeböhm, the brother of Carl Duisberg's wife, Johanna, and a nephew of Carl Rumpff's"? It is all convincing enough as research, but might seem tedious to all but specialists nearing retirement and readers with extravagant tastes for the inessential.

The story doesn't really gather steam until more than halfway through, when the focus shifts from the chronic sharp practice of the market place to the ideas produced in various centres of independent research. The leading manufacturers of aspirin play only a modest role in the discoveries of the many other useful properties of the drug made over the past 20 years. Why would even a leading manufacturer of aspirin put resources into basic research on the drug, when the fruits would be available to any other company to exploit?

Once the authors begin to discuss the work of scientists such as S. K. Bergstrom, H. Collier, R. Peto, B. Samuelson and J. Vane, *The Aspirin Wars* turns into a riveting read. Of particular interest is the recent discovery that aspirin is important in the treatment of cardiovascular disease, a finding that has made it the world's most successful drug, consumed by one in five Americans every day. By the late-1980s, aspirin was riding so high that even the manufacturers of Anacin came clean: for many years, and despite much regulatory pressure, advertisements for Anacin had never actually disclosed that the product was basically aspirin. Only when the bandwagon really began to roll did they admit it — although they had "a better aspirin formula", of course.

Nevertheless, the authors conclude with a lament: "Without a quick buck to squeeze from the new uses for aspirin, one wonders whether these benefits, discovered after so much effort by so many talented scientists, will ever reach the people most in need of them." Meanwhile, Merck Sharp & Dohme says that it no longer actively promotes cyproheptadine as an appetite stimulant, although the drug is still widely prescribed as such in some of the poorest countries in the world. □

Charles Medawar is at Social Audit Ltd, PO Box 111, London NW1 8XG, UK.

Corridors of power

Charles C. Gillispie

Science under Control: The French Academy of Sciences 1795–1914. By Maurice Crosland. Cambridge University Press: 1992. Pp. 454. £60, \$120.

It is Maurice Crosland's argument that insofar as the science done in any country partakes of a national character, development under the aegis of the state is what has given French science its distinctive style. The Academy of Sciences has been the central institution in French science since its foundation under Louis XIV in 1666. Suppressed as a privileged body by the revolutionary

that the state interfered with the actual conduct of scientific investigations or exercised any form of censorship over the findings. What was subject to control was the career pattern of scientists, not the content of theories. Sociologically oriented historians, who are currently very influential, argue that these two aspects of science, the cognitive and the institutional, cannot be separated so categorically, and some say not at all. Crosland takes it for granted that they can, however, and does not address the issue of the social construction of knowledge. His subject is thus the ways in which the Academy of Sciences shaped the practice of science in nineteenth-century France.

His is a thematic rather than a narrative approach. He treats his period, that of bourgeois France, as a block of time in the history of the Academy. The



government in 1793, it reemerged after the Terror as the First Class of the Institute of France, the other two classes, much less important at the outset, being concerned with political and social sciences and with arts and letters. Provision for the Institute, the embodiment of culture and learning, is contained in the very Constitution of the Republic drawn in 1795. Although there have been several modifications, among them revival of the term "Academy" with the Bourbon Restoration in 1815, the status of the Institute, like that of other vital organs of the French body politic, has been largely unaffected by all the subsequent changes of regime.

The phrase "under control" in Crosland's title must not be taken to imply

control that this body exercised is not to be compared to that of a Centre National de la Recherche Scientifique, nor yet of a Soviet Academy of Sciences *avant la lettre*. Nothing managerial, academic power was nonetheless effective for being exercised judgements. Election to membership continued to be the highest goal of scientific ambition throughout the century, one increasingly difficult of attainment since the number of places increased only from 60 to 78. Division of the Academy into ten, later eleven, sections advanced the professional standing of the several disciplines. It also made provision for their representation in a central forum, thus mitigating fragmentation of science among the specialized societies that proliferated in France

as elsewhere.

The weekly *Comptes rendus* of the academy, started in 1835 to supplement the more leisurely and magisterial annual volume of *Mémoires* suited to the eighteenth century, remained the vehicle of choice for publication of research even as new journals multiplied around them. Registration of the fact and date of a piece of work in a sealed note deposited with the Permanent Secretary continued to be the preferred method for safeguarding a scientist's priority in a discovery he was not ready to publish. An elaborate system of prizes permitted the Academy to point directions for research by setting the subjects for competitions and to reward the kind of work it wished to encourage by making appropriate awards. The amounts were often large, on a scale comparable to that of research grants in modern times. Finally, and most important, the Academy was the centre of patronage in the scientific community. Academicians generally held the appropriate chairs in the four citadels of science and learning, the Collège de France, the Muséum d'Histoire Naturelle, the Ecole Polytechnique and the Ecole Normale Supérieure. The Academy also had the power of nominating candidates for appointment by the Minister of Education throughout the entire university system.

Ceremonially, the place of the Academy within the Institute integrated science into culture in the eyes of the French public. At the open sessions of the Institute there were always spectators — there still are — to watch the eminent old men in their black and green Napoleonic uniforms, swords dangling awkwardly at their sides, descend the staircase and enter the stately hall under the dome to the tattoo of drummers from the Garde Républicaine. In its weekly meetings, closed to the public, the Academy received communications, heard reports and papers, and appointed commissions, attending to the business of science with a regularity and frequency unknown in Britain, the United States or Germany. Internationally, too, the Academy could represent French science with a coherence unmatched in other countries. It is no accident that the metric system was French in origin.

Crosland brings out the detail of the operations of the Academy with the meticulous clarity that is the hallmark of his scholarship. He does not make judgements about the effect on the quality, originality or range of science done in France, or enter into the question whether there was a decline in scientific vitality relative to Germany and Britain in the course of the nineteenth century. Neither in this nor in his other writings is

he given to grand views, and his quiet voice comes as rather a relief amid the extravagance of much current writing about the scientific enterprise. My only complaint about his book is that, oddly enough, it contains no bibliography and that the full and thorough footnotes are not covered by the index. The reader looking for items in the literature will have to search through them all. □

Charles C. Gillispie is emeritus professor of history of science at Princeton University, Princeton, New Jersey 08544, USA.

Pesticides in perspective

Alastair Hay

The British Medical Association Guide to Pesticides, Chemicals and Health.

Edward Arnold: 1992. Pp. 215. £16.95.

Handbook of Pesticide Toxicology: Vol. 1, General Principles; Vol. 2, Classes of Pesticides; Vol. 3, Classes of Pesticides. Edited by Wayland J. Hayes Jr and Edward R. Laws Jr. *Academic: 1991. Pp. 1,576. £250, \$425.*

Fate of Pesticides and Chemicals in the Environment. Edited by J. L. Schnoor. *Wiley: 1992. Pp. 436. £109, \$164.*

PESTICIDES are often in the headlines. In some ways, they are like the stars of stage and screen: we cannot read enough about them and the more salacious the tale the better we like it. For some sections of the industry, most of what has been written is simply ill-informed gossip. Unless pesticides are portrayed as a sort of deliverer, as molecules that swell our food baskets and kill our pests, they are not considered to be worth writing about. But this attitude is now very much the preserve of a minority.

If evidence were needed that times have changed, it is to be found in the British Medical Association's (BMA) guide to *Pesticides, Chemicals and Health*. Produced under the tutelage of a BMA working party, which included representatives from Friends of the Earth and the Transport Workers' Union, the book is an excellent introduction to a thorny subject. Such is its appeal that the organization that represents pesticide manufacturers in the United Kingdom, the British Agrochemicals Association (BAA), is now also recommending it. The BAA provided a good deal of the background literature for the working group, including a detailed assessment of the occupational health risks involved in pesticide manufacture and use.

But the BAA was not always so favourably inclined towards the BMA's

venture. A year-and-a-half ago, the trade association viewed the BMA working group's initial report more like a poisoned pen letter than a statement from the United Kingdom's principal medical body. The BAA's aversion to many of the recommendations in the report was so great that it organized a massive lobbying campaign. Relationships between the two associations became decidedly frosty. Fortunately, they have now thawed and it would seem that the BAA has had considerable second thoughts about its position. The book it is now endorsing is also a far more polished document than the earlier report, although most of the recommendations remain the same.

The volume is a lavishly illustrated guide that reviews the history of pesticides and discusses how they are formulated and tested, where they are used, what laws govern their use and how they can affect the wider environment. All this is by way of introduction to the meat of the book, which is about health risks. Drawing on evidence that the BMA collected for an earlier publication on the public perception of risk, the book points out that "public perception of risk bears little relation to the degree of risk when expressed as a mathematical probability, but depends on several other factors, such as whether the taking of risks is voluntary or involuntary, whether the hazards affect individuals or the whole of society" and whether damage is transitory or more permanent. As the guide ably illustrates, the experts and members of the general public differ markedly in their assessment of the risks of nuclear power, swimming and mountain climbing. There is relatively little disagreement, however, about the dangers of cars, motorbikes, handguns and, curiously, pesticides.

The BMA believes that the United Kingdom requires a pesticides policy that should have five key themes. These are a commitment to reduce the use of the chemicals; legislation to ensure that information about all chemicals is freely available; a policy of diversification so that natural predators of pests can be harnessed to help control infestations; a revised regulatory system to limit any conflicts of interests; and better education of doctors to recognize pesticide poisoning. Even though there is now legislation in the United Kingdom that requires pesticide sprayers to be properly trained, the BMA notes that there is still worry about the longterm health consequences of exposure to pesticides and much that is still not known about the toxicity of these chemicals. The BMA therefore concludes that the "benefit of the doubt should be awarded to protecting the environment, the worker and the consumer".

For those who want a more detailed look at individual pesticides there is probably no better source than the *Handbook of Pesticide Toxicology*, an updated version of Hayes's splendid *Pesticides Studied in Man* (Williams and Wilkins, 1982). Like its predecessor, the new handbook discusses general principles of pesticide toxicology before reviewing the evidence for the effects of the chemicals on humans, domestic animals and wildlife. It is intended as a complete guide to the field, with chapters written by acknowledged experts. Each chapter is well researched and information about particular properties of individual pesticides is easy to find. The handbook is a real treat for those who can afford it.

The wider impact of pesticides is addressed in *Fate of Pesticides and Chemicals in the Environment*, a legacy of a 1972 accord between President Richard Nixon of the United States and Leonid Brezhnev, general secretary of the Soviet Communist Party. It is the product of an international symposium held under the auspices of the US Environmental Protection Agency and involving scientists from the United States and the former Soviet Union. According to the book's editor, the fate of pesticides and chemicals formed an important bilateral agreement between the two countries.

Fate of Pesticides and Chemicals in the Environment is a very specialized book that discusses various models of the movement of chemicals in air, soil and water. Some chapters deal with processes that may affect retention in the environment, such as oxidation-reduction reactions of pollutants in the bottom sediments of lakes. Others deal with microbial transformation of chemicals in aquatic systems. The end result is a tome that is likely to appeal to a select group of environmental scientists. In common with most other books on pesticides, this one also peers into the future. It should come as no surprise that many of its predictions about the way ahead are not unlike those that the BMA is recommending. In their conclusion to the book, Julius Menn and Lawrence Christy of the US Department of Agriculture express their belief that the "agrochemical industry will continue to introduce safer, more efficacious and selective synthetic chemicals". They believe that many of these will be eventually replaced by natural-product pesticides and plants resistant to insects and disease. I hope that we will not have long to wait. □

Alastair Hay is in the Department of Chemical Pathology, University of Leeds, Old Medical School, Thoresby Place, Leeds LS2 9JT, UK.



Nigel Cattlin/Holt Studios International

Ups and downs — aerial crop spraying accounts for no more than two per cent of all pesticide applications, but is a main focus of public concern.

Realm of the senses

Wolfgang Marwan and Dieter Oesterhelt

Sensory Receptors and Signal Transduction. Edited by John L. Spudis and Birgit H. Satir. Wiley: 1992. Pp. 269. £79, \$118.

It has been known for more than a hundred years that cells, including unicellular organisms, can sense their environment and respond to it in a remarkable variety of ways. The molecular mechanisms underlying this basic phenomenon of signal transduction are now beginning to be unravelled. Chemical compounds, light or other physical stimuli are sensed by receptor proteins, which, in the process, undergo a conformational change. The conformational signal triggers a cascade of events that ultimately alter cell behaviour. In many systems, the sensory machinery is highly dynamic: the sensory input may be amplified by catalytic cascades; the information from several types of receptors may be integrated; and feedback loops may mediate adaptation to the stimulus, allowing the detection of temporal changes in receptor excitation.

The main focus of this volume is the biochemical processes involved in photo- and chemosensing in cells of organisms ranging from bacteria to mammals. The five articles cover a vast range of phenomena, including archaebacterial phototaxis, phytochrome-regulated plant development, neutrophil leukocyte che-

motaxis and vertebrate olfaction. The book is appealing because it provides a more or less systematic overview of cellular sensing, combined with a more detailed discussion of some well-understood model systems.

The systematic tour through the various types of photoreceptors requires comparative discussions of systems such as vertebrate vision, which is well understood, and blue-light-modulated gene expression in plants, which is not. This makes the book gripping to read, and immediately highlights the gaps in our knowledge. The same is true for the coverage of chemoreception. The chapter on bacterial chemotaxis gives a detailed review of the enormous knowledge that has accumulated over the past 20 years. This section is by far the most extensive in the book, in keeping with the current status of the enterobacterial chemotaxis system as the best-understood signal-transduction network.

The book is clearly written and easily understandable throughout. It provides a valuable point of access to original literature, with references up to 1991 for some chapters. The volume will be useful to all researchers working on signal transduction or in photobiology. Graduate students interested in the subject, which is barely covered by standard textbooks on biochemistry or biology, will also profit from reading it. And, as a fine example of the state of the art, it will be sure to inspire newcomers to the field of sensory reception and signal transduction. □

Wolfgang Marwan and Dieter Oesterhelt are in the Max-Planck-Institut für Biochemie, Am Klopferspitz 18A, 8033 Martinsried, Germany.

Rhizobium–plant signal exchange

Robert F. Fisher & Sharon R. Long

Initial stages in the *Rhizobium*–legume symbiosis can be thought of as a reciprocal molecular conversation: transmission of a gene inducer from legume host to bacterium, with ensuing bacterial synthesis of a morphogen that is transmitted to the plant, switching the developmental fate of the legume root. These signal molecules have a key role in determining bacterium–host specificity and the purified Nod factor compounds provide useful new tools to probe plant cell function.

THE *Rhizobium*–legume symbiosis culminates in the formation of nitrogen-fixing root nodules. These unique structures are agronomically significant, as they provide an alternative to the use of energy-expensive ammonium fertilizer. As a system for the study of developmental biology, *Rhizobium*–root nodule formation provides unique experimental opportunities: each partner can be manipulated so as to permit genetic and molecular dissection of the developmental sequence¹.

Several genera in the *Rhizobiaceae* nodulate legumes: *Rhizobium*, *Bradyrhizobium* and *Azorhizobium*². Here we will use the term rhizobia for all three genera. Individual bacterial species and strains nodulate a particular set of host plants (Table 1), and are characterized as having a host range that is either broad (nodulating many different plants) or narrow (nodulating one or few hosts)². Nodules formed on different plants by different bacteria nonetheless display striking developmental similarities (for an outline of early nodulation events, see Fig. 1). Thus, there are host-specific triggers for essentially homologous processes in different hosts: this has been the basis for useful experimental approaches to the dissection of the early nodulation process^{3–5}.

Recent results show that the initial nodulation steps can be characterized as a two-way molecular conversation. The host legume releases signal compounds that stimulate the coordinate expression of bacterial genes that are required for nodulation (*nod* genes). These *nod* genes, in turn, encode enzymes involved in the synthesis of Nod factors: substituted oligosaccharides that cause morphological changes in the plant root. These Nod factor compounds act as determinants for the host range exhibited by a specific bacterium. Echoing the structural similarity of nodules in different host–bacterial systems, the molecular signals in the rhizobia–legume symbioses are variations on a theme. Inducers from the plant are phenolics of the flavonoid group, with skeletal or side group modifications to distinguish the inducers from distinct plants; the morphogens synthesized by the bacteria are all oligomers of *N*-acetyl glucosamine, with different host-specific modifications. We will review here what is known about this two-way molecular signalling between plant and bacterium, and discuss current theories on the roles that these signal compounds have in gene regulation and plant development.

TABLE 1 *Rhizobium* nodulation host range

Bacterial species	Plant hosts
<i>Rhizobium meliloti</i>	alfalfa (<i>Medicago sativa</i>), <i>Melilotus albus</i>
<i>Rhizobium leguminosarum</i>	
biovar <i>viciae</i>	pea (<i>Pisum sativum</i>), vetch (<i>Vicia sativa</i>)
biovar <i>trifolii</i>	clovers (<i>Trifolium</i> species)
biovar <i>phaseoli</i>	<i>Phaseolus</i> bean
<i>Bradyrhizobium japonicum</i>	soybean (<i>Glycine max</i>)
<i>Azorhizobium caulinodans</i>	<i>Sesbania</i> (root and stem nodules)
<i>Rhizobium</i> spp. NGR234	broad host range; genera including <i>Vigna</i> , <i>Macroptilium</i> , <i>Lablab</i> , <i>Glycine</i>

Representative bacterial symbionts and some of the plants that they nodulate effectively. This list is not complete, as it shows only those species referred to in this review; other bacterial species have been described, and the taxonomic classifications are constantly being rediscussed. The bacteria listed here in many cases have additional hosts to those listed.

Host range is controlled by *nod* genes

The *nod* genes (Fig. 2) have been characterized as either 'common', such as *nodABC*, which are essential for nodulation to occur^{2,6,7}, or 'host-specific', which are found only in certain strains and help to determine the host range shown by the bacterium. Examples of host-specific genes are *R. leguminosarum nodE*, which specifies nodulation of peas or clover⁸ depending on the biovar source of the gene (*viciae* or *trifolii*, respectively), and *R. meliloti nodH* and *nodPQ*, which are responsible for specifying nodulation of alfalfa^{9,10}. Generalizations about common and host-specific *nod* genes may be challenged and extended by broad host-range bacteria such as NGR234, where host range may be 'modular', that is, determined by several genetically dispersed sets of host-range genes^{11–14}. The *nodD* gene is expressed in free-living cells; other *nod* genes are induced in the presence of the plant host^{1,2,6}, a behaviour similar to that of the closely related *Agrobacterium tumefaciens*, whose *vir* genes are induced by compounds released from plant wounds¹⁷. The molecules that induce the rhizobial *nod* genes are flavonoids^{18,19}, which are synthesized as products of the central phenylpropanoid pathway in plants²⁰, and a few other

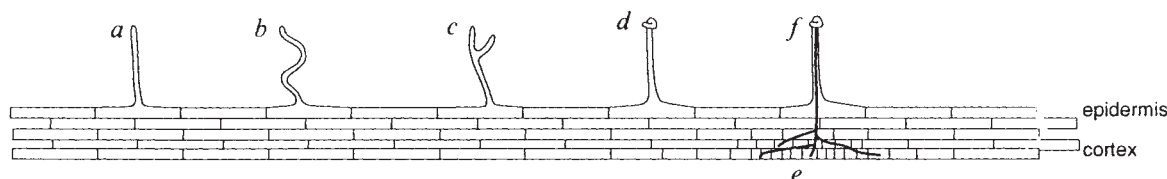


FIG. 1 Initial stages in the rhizobia–legume symbiosis. Rhizobial signalling to host root hairs (a) leads not only to root hair deformation (b), branching (c) and curling (d), but also to mitosis in the root cortex (e), which culminates in the formation of the nodule primordium. Bacteria invade the plant cells

through a novel structure termed the infection thread (f), a plant-derived tube which delivers the bacteria into individual cells within the nodule primordium. At this point the bacteria differentiate into bacteroids, which can fix atmospheric nitrogen into ammonia.

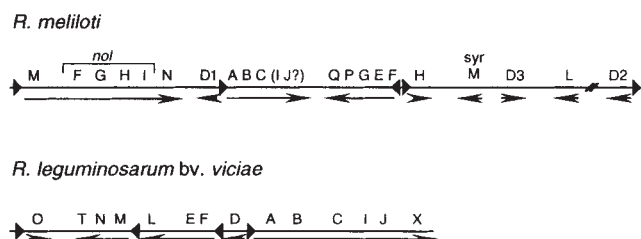


FIG. 2 Map of *R. meliloti* and *R. leguminosarum* bv. *viciae* nod genes. Arrows indicate orientation of nod gene expression, and triangles indicate location and orientation of nod boxes¹⁵. As shown in complementation studies, DNA segments representing what is likely to be a complete set of nod genes have also been identified for *R. leguminosarum* bv. *trifolii*¹⁶. Many nod genes have also been identified in other species, including *R. leguminosarum* bv. *phaseoli*, *Rhizobium* sp. NGR234, *B. japonicum*, *A. caulinodans* and *R. loti*. The nodABC and the regulatory nodD are found in all rhizobia examined; the genetic arrangement is in many cases dispersed, making it a challenge to identify a complete set of genes. Among the genes in these other species are probably a number that encode host specificity determinants.

phenolics²¹. Different plants secrete different flavonoid inducers^{18,19,22-25} and a single plant may make different inducers at different times during development²⁶. Besides their role in nod gene induction, some flavonoids cause growth enhancement of *Rhizobium*; thus these compounds may have a significant and as yet unexplored role in the overall soil ecology of microbes¹⁹.

nod gene regulation

Plant inducer-mediated expression of nod genes such as nodABC requires NodD (Fig. 3)²⁷⁻²⁹. The nodD gene exists as a single gene in some rhizobia, but as a multi-gene family in others³⁰⁻³³. The nodDs from diverse sources differ in their response to various hosts, behaving in a species-specific manner; they are generally most active in combination with extracts from plants that they normally nodulate³⁴⁻³⁷. Perhaps the different multiple alleles of nodD in certain species allow the sensing of different host-specific inducers which then optimize nodulation by symbiotic partners^{26,32,38}. The N-terminal halves of NodD proteins are much more highly conserved than the C-terminal halves^{31,34,39}, implying that the C-terminal portion delineates a variable function such as binding to different flavonoids, whereas the N-terminal portion specifies DNA binding²⁸. Analysis of mutant and chimaeric NodDs suggests that flavonoid specificity is largely determined by the C-terminal portion of NodD^{34,40,41}. But the overall tertiary structure probably also contributes to the formation of functional domains⁴² (DNA-binding, flavonoid interaction and activation).

Most nod operons are preceded by a highly conserved sequence termed a nod box¹⁵ (Fig. 3). We found that NodD binds to two sites along one face of the nod box^{43,44}, in contrast to a proposal that important sequences in the *B. japonicum* nod box lie in four smaller direct repeats on different faces of the helix⁴⁵. Moreover, NodD induces a bend in the DNA linking the two binding sites⁴⁴; the function of bending in the mechanism of nod gene activation, the nature of the NodD-RNA polymerase interaction, and the role of plant signal are unknown. How intricate might these interactions be? Mutation of *R. meliloti* groEL prevents expression of nodD3 and reduces the activity of NodD1 and NodD3 when they are expressed constitutively (ref. 43 and J. Ogawa and S.R.L., manuscript in preparation). GroEL, a 'molecular chaperone', may assist the formation of active NodD molecules or promote assembly of NodD with RNA polymerase or other unidentified components of the transcription apparatus.

Control of nodD expression is highly diverse: some nodD genes are unregulated²⁷, others are subject to negative

autogenous regulation²⁹, to repression by NodR⁴⁶, or to inducible expression⁴⁷. SyrM, which like NodD is a member of the LysR family of positive transcriptional activators⁴⁸, regulates one *R. meliloti* nodD gene through non-nod box promoters, as do some NodDs^{39,99,101}. The functions of NodDs may also be diverse, raising the question as to why some strains have multiple nodDs. Explanations citing roles in host range or developmental flexibility have been proffered; in the most-studied case (*R. meliloti*) nodD participates in both. The two inducer-requiring nodDs (nodD1 and nodD2) are important for optimal responses to different hosts^{32,34}. The inducer-independent circuit controlled by nodD3⁴⁹ may be important for late nod gene expression^{50,99,101} and/or may respond to nitrogen, growth and/or inducers⁵¹⁻⁵³. The complexities found so far in *R. meliloti* suggest that diversity of signals and strategies for nod gene induction in other rhizobia will lead to better understanding of nodule physiology and development.

Action of nod genes

A search began more than 20 years ago for the factor that completes the first round of dialogue between the endosymbiont and host, triggering plant morphogenic changes that lead to formation of nitrogen-fixing root nodules. Yao and Vincent⁵⁴ found that bacteria-free filtrates of rhizobia induced root hair deformations (Had phenotype; see Fig. 1) that paralleled the host range exhibited by intact cells. By correlating bioassay activity (such as Had or root shortening) with nod genotype and nod gene expression, the connection between genetics and filtrate properties was established^{3,55}. Filtrates from Nod⁺ luteolin-induced cultures of *R. meliloti* carrying extra copies of the nod genes produced high levels of alfalfa Had-active factor⁴. Organic extraction and chromatography coupled with Had bioassays resulted in purification of NodRm-IV(S) (originally NodRm-1), a sulphated and acylated β -linked *N*-acetylglucosamine tetrasaccharide (Fig. 4). Correlation of NodRm-IV(S) with nod genes was demonstrated by Tn5-inactivation of nodA or nodC, which abolishes both production of the factor molecule and filtrate bioassay activity, strongly suggesting that nod genes encode enzymes that synthesize the factor. Furthermore, this newly described oligosaccharide seems to be the nodule morphogen, as described below.

The basic structure of Nod factors from several organisms seems to be a β -1,4-linked oligomer of *N*-acetylglucosamine,

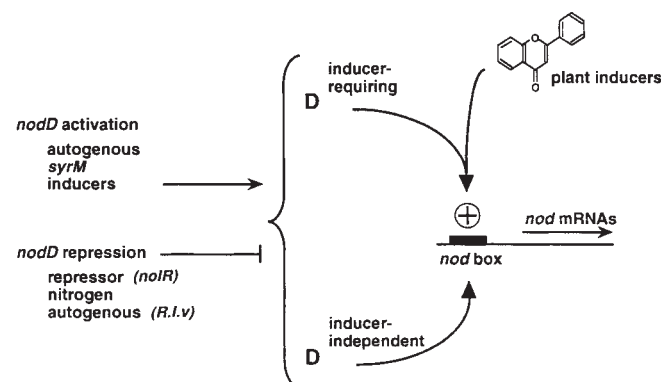


FIG. 3 Regulatory circuitry of nod gene expression. Inducible nod genes can be activated either by inducer-requiring forms of NodD in conjunction with plant flavonoids^{27,29,47,84-86}, or by inducer-independent forms of NodD in the absence of any host components^{38,49}. NodD binds to nod boxes upstream of the inducible nod genes^{46,87,88}. In *R. meliloti*, syrM and nodD3 (Fig. 2) participate in a self-amplifying cascade leading to high levels of expression of nodD3^{99,101}. In some rhizobia expression of nodD is induced by plant flavonoids; such induction requires NodD^{47,30}. Expression of inducible nod genes can be indirectly controlled by autogenous regulation of nodD²⁹, and also by regulators such as NodR⁴⁶, which represses expression of the activator nodD genes.

having an *N*-acyl substitution on the non-reducing end (Fig. 4). Two further generalizations emerge^{8,56,57}: first, individual bacterial strains make a family of factors that may differ slightly in length and/or substitution; second, substitutions usually differ when factors from different species are compared. Thus factor structure may account not only for the major host range distinctions between species and biovars of rhizobia, but also for the 'fine-structure' host range seen in the preferential or optimal nodulation of some plants in a cross-inoculation group by some strains but not others of the corresponding bacterial set. This leads to the prediction that effects of *nod* genes on factors and factor structure should parallel their effects on phenotypes of living bacteria.

Nod factor synthesis and *nod* genes

Attention is currently focused on the roles that individual *nod* genes have in the synthesis of the plant morphogenic factors (Fig. 5). One example correlates genes and enzymes with the NodRm-IV(S) sulphate substitution and with host range. Faucher *et al.*³ showed that *R. meliloti* genes *nodH* and *nodQ* determine the changes in host range observed both for bacterial cells and for cell filtrates. Our laboratory found that NodP and NodQ, homologues of *Escherichia coli* CysD, CysN and CysC, encode ATP sulphurylase and APS kinase (ref. 58 and J. Schwedock and S.R.L., unpublished results) and thus are involved in production of activated sulphate (PAPS), which

could then be transferred to NodRm-IV precursors. Indeed, Roche *et al.*⁵⁶ have now shown that mutations in either *nodH* or *nodQ* lead to production of an unsulphated factor which is no longer active on alfalfa, *R. meliloti*'s usual host plant, but is active on a different plant, vetch (*Vicia sativa*). Thus the genes, and some of the enzymes, responsible for sulphate modification have been established and correlated with a role in host range.

Host range is also specified at the other end of the Nod factor molecule. NodF and NodE, homologous to *E. coli* acyl carrier protein²⁸ and β -ketoacyl synthase (condensing enzyme)¹⁰⁰, respectively, are required for addition of specific fatty acyl chains to the non-reducing end of the signal compounds⁸. Other biochemical activities are inferred from homologies, structure of mutant factors, and direct assay. Table 2 summarizes our current knowledge and hypotheses about the roles that many of the *nod* genes have. Of particular interest is the homology of NodC to chitin and cellulose synthases (ref. 59, and E. M. Atkinson and S.R.L., and F. DeBelle and J. Denarié, manuscripts submitted), which suggests it has a role in the growth of the oligosaccharide backbone. Analysis of Nod factors from *nodL* mutants suggests that NodL adds the *O*-acetyl group to the NodRlv factor⁸. NodM is involved in glucosamine-6-phosphate synthesis, and it might provide precursors for synthesis of Nod factors^{60,61}. NodG shows homology to reductases⁶²⁻⁶⁴, and may be involved in modification of the fatty acyl side chain. NodI together with NodJ may make up part of a bacterial membrane transport

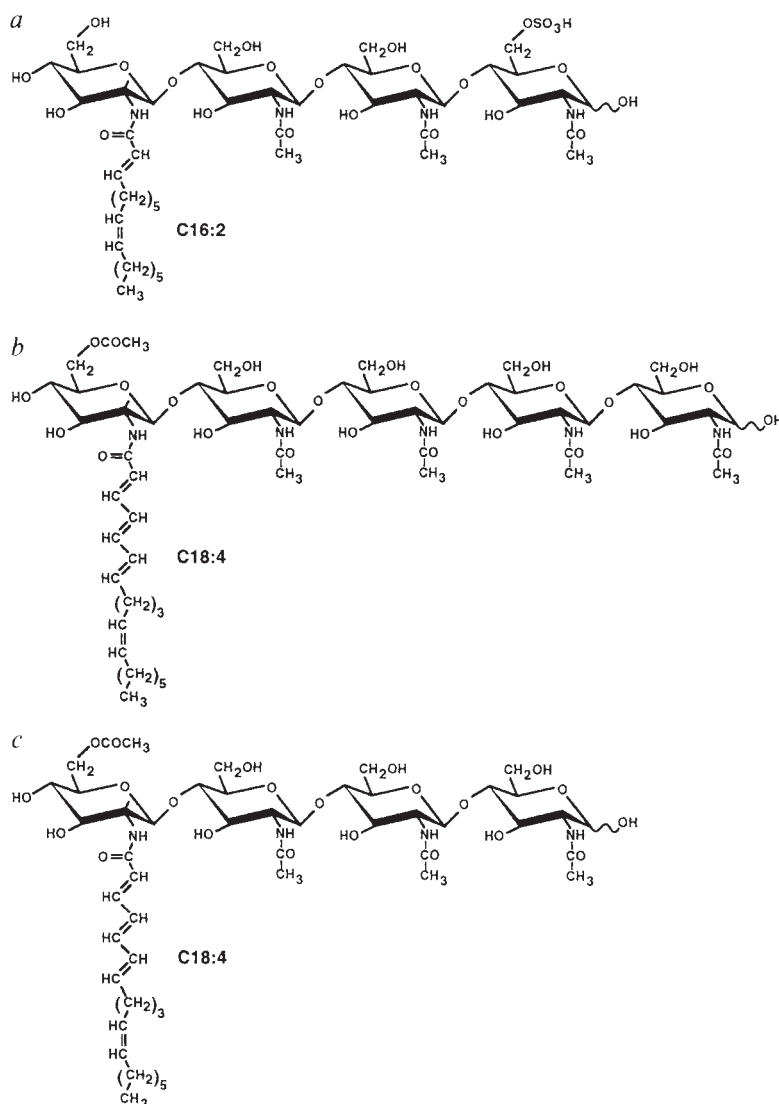


FIG. 4 Structures of bacterial Nod factors that induce nodule morphogenesis. The three Nod factors shown here illustrate common and variable features of this family of molecules. *a*, NodRm-IV(S)⁴; *b*, NodRlv-V(Ac)⁸; *c*, NodRlv-IV(Ac)⁸. Nod factors from different bacterial sources differ in their fatty acyl substitutions (compare C16:2 of the *R. meliloti* Nod factor *a* with C18:4 of the *R. leguminosarum* bv. *viciae* Nod factor (*b* and *c*)). Nod factor backbone length can also vary: *b* and *c* are wild-type pentameric and tetrameric versions, respectively, of Nod factor from *R. leguminosarum* bv. *viciae*. New Nod factors with other modifications are being identified in a number of rhizobial systems.

complex^{65,66}. The Ca²⁺-binding NodO is proposed to interact with plant roots and to mediate an early stage in rhizobia-legume recognition⁶⁷.

The finding of multiple factors within species raises questions, and provides clues, about synthetic mechanism and symbiotic significance. For example, cells produce both tetramer and pentamer variants of Nod factors^{8,56}, raising questions on the biogenesis of factors of differing backbone length. Are glucosamine subunits added sequentially, or are long glucosamine polymers cleaved into tetra- and pentasaccharide units and subsequently modified? What specific role different Nod factor variants have in the symbiotic interaction is unknown. They may be a manifestation of the host range displayed by the individual *Rhizobium* species; that is, each variant factor may be fine-tuned to interact optimally with one of the plants that it nodulates. Moreover, there may also be as yet undetermined environmental factors that shift the spectrum of factors which are made at any given time.

Much research is focused on the biogenesis of these signal molecules⁶⁸. Ultimately it is necessary to synthesize active factors *in vitro* using purified gene products as enzymes to confirm their biosynthetic pathways. This is critical for two reasons: first, to preclude the possibility that the plant morphogenic activities observed are the result of the presence of low level, structurally distinct Nod factor compounds co-purified from extracts;

second, to achieve the ultimate demonstration that *nod* genes encode the actual enzymes for synthesis and are not merely the regulators for enzymes encoded elsewhere.

Plant response

Rhizobial cells elicit multiple host reactions, but are the numerous plant responses caused by equally numerous stimuli, or by only one or a few signals from the bacterium? Purified Nod factors allow this to be tested. NodRm-IV(S) and an acetylated form (NodRm-IV(Ac,S)) elicit both cortical cell divisions and nodule formation on alfalfa seedlings⁵, satisfying the requirements of the nodule organogenesis-inducing principle postulated over a decade ago⁶⁹. Nod factors also stimulate plant expression of 'early nodulins', host proteins expressed during the initial response to rhizobial infection. Thus the same factors can cause both root hair deformations and nodule organogenesis. It will be important to test whether different plant responses are maximally triggered by the same, or slightly different, Nod factors produced by any one bacterium (T. Bisseling, personal communication). On *Vicia*, for example, non-specific Nod factors cause root hair deformation, but fully host-specific factors cause both root hair deformation and cell division⁸. Such information provides suggestions for, and imposes constraints on, models for Nod factor receptor action.

With what plant systems do Nod factors interact? Nodule morphogenesis, including early nodulin gene expression, is mimicked by auxin transport inhibitors and by cytokinin⁷⁰⁻⁷². Could nodule growth be caused by changes in root auxin or cytokinin levels, transport, or sensitivity? Or do these pathways intersect in some indirect way? Finally, we need to know with what components of the growing root hair (cytoskeleton, membrane, cell wall) do Nod factors interact to cause root hair deformation?

The structural information present in the Nod factors has yet to be completely decoded, but both ends of the molecule are important^{8,56}. Is there just one receptor in any one plant for all of the factors secreted by its particular symbiont? If so, how does such a receptor recognize both ends of a molecule, while tolerating more than one overall length of the ligand? In this light, we may need to envision the existence of a family of Nod factor receptors. Models to explain receptor responsiveness are subject to several initial constraints. One is that the acyl tail of the known Nod factors will require a substantial hydrophobic cleft in the protein, or possibly will require interaction with the plant membrane lipids or with a protein domain buried in the

TABLE 2 Functions and properties of *nod* genes

<i>nod</i>	Known function or properties
AB	Required for Nod factor production ⁴
C	Homology to chitin and cellulose synthases ⁵⁸ ; proposed to form β -1,4-glycosyl bond
D	Transcriptional activator of inducible <i>nod</i> genes ²⁷⁻²⁹
E	Host-specific; homology to β -ketoacyl synthase (condensing enzyme) of fatty acyl synthase ¹⁰⁰ ; proposed to synthesize Nod factor acyl chain
F	Host-specific; homology to acyl carrier protein ²⁸ ; proposed to synthesize Nod factor fatty acyl chain
G	Host-specific; homology to reductases ⁶²⁻⁶⁴ ; proposed to modify Nod factor fatty acyl side chain
H	Host-specific; required for formation of sulphated Nod factor ⁵⁶ ; proposed to transfer activated sulphate (PAPS) to Nod factor
I	Homology to ATP-binding active transport proteins ^{65,66} ; proposed to form membrane transport complex with NodJ
J	Homology to transmembrane proteins ⁶⁵ ; proposed to form membrane transport complex with NodI
L	Host-specific; homology to acetyl transferase ⁸⁹ ; required for formation of 6-O-acetyl Nod factor ⁸ ; proposed to add O-acetyl group to Nod factor
M	Host-specific glucosamine synthase ⁶⁰ ; proposed to synthesize Nod factor sugar subunits
N	Host-specific; involved in <i>Vicia hirsuta</i> nodulation ⁹⁰
O	Exported Ca ²⁺ -binding ⁶⁷ ; homology to haemolysin; proposed to mediate early stage in rhizobia-legume interaction
P	Host-specific ATP sulphurylase ⁵⁸ ; proposed to provide activated sulphate for transfer to Nod factor
Q	Host-specific ATP sulphurylase and APS kinase ⁵⁸ ; together with NodP makes activated sulphate (PAPS); proposed to provide activated sulphate for transfer to Nod factor
T	Host-specific; involved in <i>Trifolium subterranean</i> nodulation ⁹¹ ; proposed to be membrane protein
V	Homology to two-component regulatory system sensor proteins ⁹² ; proposed to regulate unknown target genes
W	Homology to two-component regulatory system activator proteins ⁹² ; proposed to regulate unknown target genes
X	Host-specific hydrophobic protein; extends host-range to Afghanistan peas ⁹³
<i>nodR</i>	Repressor of <i>nodD</i> ⁴⁶

Other genes including *nodK*⁹⁴, *nodS* and *nodU*¹⁴, *nodY*⁹⁵, *nodZ*⁹⁶, *nodA*⁹⁷, *nodE* and *nodF*⁹⁸, and *nodI*, *nodG*, *nodH* and *nodI*⁶⁰ have been identified, but have not been described by means of sequence homology to other published gene products, nor have possible functions been proposed.

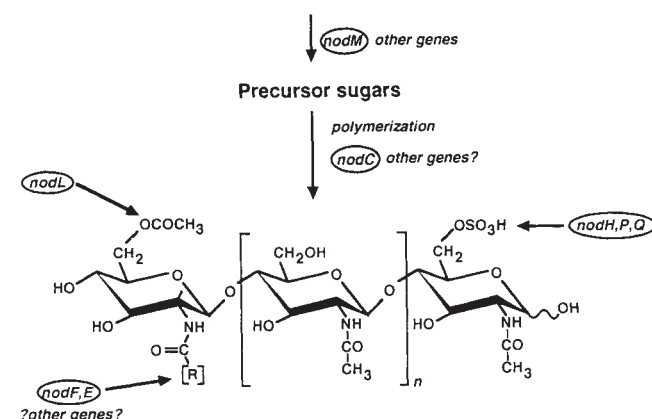


FIG. 5 Hypothetical pathway of Nod factor biogenesis. Nod factor sugar precursors, proposed to be synthesized by the glucosamine synthase activity of NodM⁶⁰, could be polymerized by the action of NodC, which is homologous to chitin and cellulose synthases (ref. 59 and E. M. Atkinson and S.R.L., manuscript in preparation). Strains bearing mutations in *nodFE* yield Nod factors with altered fatty acyl chains¹⁰. Other Nod factor modifications seen in different bacteria include the C-6 sulphate (Fig. 4a) added by action of *nodH* and *nodPQ* in *R. meliloti*^{56,58} and the C-6 acetyl group that requires *nodL* in *R. leguminosarum* bv. *viciae* (Fig. 4b and c)¹⁰.

plant membrane bilayer. Another is that each legume should have at least one receptor for the Nod factors, and that between species we may expect the receptor to have most of the same domains (those connecting the receptor to the inner workings of the plant cell, or that bind to the *N*-acetylglucosamine backbone) while having other domains that vary to recognize the modifications specific to the Nod factor tailored to that specific host. Thus there may exist a family of receptor genes in the legume family with constant and variable domains analogous to those of diverse animal receptors for immune determinants, hormones or odors^{73,74}.

Evolutionary considerations

The nature of the Nod factor receptor may instruct us about the evolution of the mutualistic symbiosis itself. The relationship between the rhizobial symbiosis and a controlled pathogenesis has been noted, as has the defence-like response of plants to failed symbioses⁷⁵⁻⁷⁸. Might the Nod factor receptor be evolutionarily related to receptors that mediate certain host defences in plants? It would present a pleasing symmetry of a sort: the signals that legumes send to rhizobial partners are flavonoids, a class of compounds which also includes the phytoalexin defence molecules. Interestingly, simple isoflavone inducers of soybean not only cause *B. japonicum nod* gene expression, they also trigger bacterial resistance to the pterocarpan isoflavonoids that soybeans produce as a phytoalexin⁷⁸. The signal that the bacterium replies with in turn structurally resembles elicitor compounds, in that chitin (unmodified β -1,4-linked *N*-acetyl glucosamine) is a signal that plants interpret as evidence of a pathogen⁷⁹. Perhaps it is the additional modifications (acyl, acetyl, sulphate and others yet to be found) that coax the plant into accepting rhizobia as a symbiont, instead of mounting a defensive response to the invader.

An alternative and not mutually exclusive proposal is that Nod factors imitate native plant growth substances, for which other similar receptors exist. For *A. tumefaciens*, the tumour-inducing mechanism relies on engineered production of known plant hormones, auxins and cytokinins⁸⁰; its generic cousin *A. rhizogenes* affects the threshold of plant response to those same hormones⁸¹. Nod factors have not been recognized as native plant growth regulators, but now that their structures are known, a search for comparable compounds in plants themselves could be initiated. One opportunity for testing this idea arises with certain alfalfa lines that generate nodules spontaneously⁸²: do such lines produce high levels of endogenous compounds similar to Nod factors; or do they have secondary signal systems that 'fire' spontaneously?

It is likely that the identification of the Nod factors, with their highly specific tailored ends for different plants, and with their ability to provoke different cell-specific responses in their host plants, will provide a tremendous impetus to research in plant cell biology. As root hairs are typically single cells, optically almost transparent, and accessible to micro-manipulation, they make attractive targets for study of how plant cells respond to exogenous chemical signals. We have found that transmembrane potentials of single alfalfa root hairs depolarize within minutes of Nod factor application⁸³. This may help studies on whether Nod factors trigger ion flux and secondary signals not only in root hair cells, but also in other cells or protoplasts whose morphology does not permit display of a classic 'root hair' response.

At many levels the study of the rhizobia-legume symbiosis has provided insight into the basic biology of both bacteria and plants, in addition to elucidating the details of the symbiosis itself. As presented here, these advances include more detailed knowledge of a class of bacterial activator proteins; insights into the function and possibly the evolution of the flavonoid compounds in plants; and discovery of a novel compound used as a plant signal, whose synthetic pathway in bacteria was previously undescribed, and the study of which is already

advanced in terms of our knowing which genes carry out certain steps. It is an illustration, perhaps, of the maxim that one often best discovers how a process or entity works by perturbing it. In the interference of rhizobia into the affairs of plants, there is an exciting history still to be read. □

R. F. Fisher and S. R. Long are at the Department of Biological Sciences, Stanford University, Stanford, California 94305-5020, USA.

- Long, S. R. *Cell* **56**, 203-214 (1989).
- Young, J. P. W. & Johnston, A. W. B. *Trends Ecol. Evol.* **4**, 331-349 (1989).
- Faucher, C., Camut, S., Denarie, J. & Truchet, G. *Molec. Pl.-Microbe Interac.* **2**, 291-300 (1989).
- Lerouge, P. et al. *Nature* **344**, 781-784 (1990).
- Truchet, G. et al. *Nature* **351**, 670-673 (1991).
- de Bruijn, F. J. & Downie, J. A. *Curr. Biol.* **2**, 184-192 (1991).
- Long, S. R. in *Biological Nitrogen Fixation* (eds Stacey, G., Burris, R. H. & Evans, H. J.) 560-597 (Chapman and Hall, New York, 1992).
- Spaark, H. P. et al. *Nature* **354**, 125-130 (1991).
- Cervantes, E. et al. *Molec. Microbiol.* **3**, 745-755 (1989).
- Schwedock, J. & Long, S. R. *Molec. Pl.-Microbe Interac.* **2**, 181-194 (1989).
- Lewin, A. et al. *Pl. molec. Biol.* **8**, 447-459 (1987).
- Lewin, A., Cervantes, E., Hee-Hoong, W. & Broughton, W. J. *Molec. Pl.-Microbe Interac.* **3**, 317-326 (1990).
- Nayudu, M. & Rolfe, B. G. *Molec. Gen. Genet.* **206**, 326-337 (1987).
- Gottfert, M., Hitz, S. & Hennecke, H. *Molec. Pl.-Microbe Interac.* **3**, 308-316 (1990).
- Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, A. & Kondorosi, A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1757-1761 (1986).
- Djordjevic, M. A., Schofield, P. R. & Rolfe, B. G. *Molec. Gen. Genet.* **200**, 463-471 (1985).
- Stachel, S. E., Nester, E. W. & Zambryski, P. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 379-383 (1986).
- Rolfe, B. G. *Biofactors* **1**, 3-10 (1988).
- Phillips, D. A. *Rec. Adv. Phytochem.* **26**, 1-33 (1991).
- Harborne, J. B. in *Secondary Plant Products* (eds Bell, E. A. & Charlwood, B. V.) 329-402 (Springer, Berlin, 1980).
- Kape, R., Parniske, M. & Werner, D. *Appl. envir. Microbiol.* **57**, 316-319 (1991).
- Peters, N. K., Frost, J. W. & Long, S. R. *Science* **233**, 917-1008 (1986).
- Redmond, J. W. et al. *Nature* **323**, 632-634 (1986).
- Zaat, S. A. J., Wijffelman, C. A., Mulders, I. H. M., Van Brussel, A. A. N. & Lugtenberg, B. J. J. *Pl. Physiol.* **86**, 1298-1303 (1988).
- Gyorgypal, Z., Kondorosi, E. & Kondorosi, A. *Molec. Pl.-Microbe Interac.* **4**, 356-364 (1991).
- Hartwig, U. A., Maxwell, C. A., Joseph, C. M. & Phillips, D. A. *Pl. Physiol.* **92**, 116-122 (1990).
- Mulligan, J. T. & Long, S. R. *Proc. natn. Acad. Sci. U.S.A.* **82**, 6609-6613 (1985).
- Shearman, C. A., Rossen, L., Johnston, A. W. B. & Downie, J. A. *EMBO J.* **5**, 647-652 (1986).
- Rossen, L., Shearman, C. A., Johnston, A. W. B. & Downie, J. A. *EMBO J.* **4**, 3369-3373 (1985).
- Davis, E. O. & Johnston, A. W. B. *Molec. Microbiol.* **4**, 933-941 (1990).
- Gottfert, M. et al. *J. molec. Biol.* **191**, 411-420 (1986).
- Honma, M. & Ausubel, F. M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 8558-8562 (1987).
- Appelbaum, E. R., Thompson, D. V., Idler, K. & Chartrain, N. J. *Bact.* **170**, 12-20 (1988).
- Horvath, B., Bachem, C. W. B., Schell, J. & Kondorosi, A. *EMBO J.* **6**, 841-848 (1987).
- Honma, M. A., Asomaning, M. & Ausubel, F. M. *J. Bact.* **172**, 901-911 (1990).
- Mclver, J., Djordjevic, M. A., Weinman, J. J., Bender, G. L. & Rolfe, B. G. *Molec. Pl.-Microbe Interac.* **2**, 97-106 (1989).
- Spaark, H. P., Wijffelman, C. A., Pees, E., Okker, R. J. H. & Lugtenberg, B. J. J. *Nature* **328**, 337-340 (1987).
- Gyorgypal, Z., Iyer, N. & Kondorosi, A. *Molec. Gen. Genet.* **212**, 85-92 (1988).
- Rushing, B., Yelton, M. M. & Long, S. R. *Nucleic Acids Res.* **19**, 921-927 (1990).
- Burn, J., Rossen, L. & Johnston, A. W. B. *Genes Dev.* **1**, 456-464 (1987).
- Burn, J. E., Hamilton, W. D., Wootton, J. C. & Johnston, A. W. B. *Molec. Microbiol.* **3**, 1567-1577 (1989).
- Spaark, G. P., Wijffelman, C. A., Okker, R. J. H. & Lugtenberg, B. E. J. *Pl. molec. Biol.* **12**, 59-73 (1989).
- Long, S. R. et al. in *Advances in Molecular Genetics of Plant-Microbe Interactions* (eds Hennecke, H. & Verma, D. P. S.) 127-133 (Kluwer, Dordrecht, 1991).
- Fisher, R. F. & Long, S. R. *J. molec. Biol.* submitted (1992).
- Wang, S.-P. & Stacey, G. *J. Bact.* **173**, 3356-3365 (1991).
- Kondorosi, E. et al. *EMBO J.* **8**, 1331-1340 (1989).
- Banfalvi, Z., Nieuwkoop, A., Schell, M., Besl, L. & Stacey, G. *Molec. Gen. Genet.* **214**, 420-424 (1988).
- Barnett, M. J. & Long, S. R. *J. Bact.* **172**, 3695-3700 (1990).
- Mulligan, J. T. & Long, S. R. *Genetics* **122**, 7-18 (1989).
- Sharma, S. B. & Signer, E. R. *Genes Dev.* **4**, 344-356 (1990).
- Dusha, I., Bakos, A., Kondorosi, A., de Bruijn, F. J. & Schell, J. *Molec. Gen. Genet.* **219**, 89-96 (1989).
- Maillet, F., Debelle, F. & Denarie, J. *Molec. Microbiol.* **4**, 1-10 (1990).
- Wang, S.-P. & Stacey, G. *Molec. Gen. Genet.* **223**, 329-331 (1990).
- Yao, P. Y. & Vincent, J. M. *Aust. J. Biol. Sci.* **52**, 413-423 (1969).
- Van Brussel, A. A. N. et al. *J. Bact.* **165**, 517-522 (1986).
- Roche, P. et al. *Cell* **67**, 1131-1143 (1991).
- Schultze, M. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 192-196 (1992).
- Schwedock, J. & Long, S. R. *Nature* **348**, 644-647 (1990).
- Bulawa, C. & Wasco, W. *Nature* **353**, 710 (1991).
- Baev, N., Endre, G., Petrovics, G., Banfalvi, Z. & Kondorosi, A. *Molec. Gen. Genet.* **228**, 113-124 (1991).
- Mane, C., Barny, M.-A. & Downie, J. A. *Molec. Microbiol.* **6**, 843-851 (1992).
- Debelle, F. & Sharma, S. B. *Nucleic Acids Res.* **14**, 7453-7472 (1986).
- Fisher, R. F., Swanson, J., Mulligan, J. T. & Long, S. R. *Genetics* **117**, 191-201 (1987).
- Baker, M. E. *Molec. Endocr.* **3**, 881-884 (1989).
- Evans, I. J. & Downie, J. A. *Gene* **43**, 95-102 (1986).
- Higgins, C. F. et al. *Nature* **323**, 448-450 (1986).
- Economou, A., Hamilton, W. D. O., Johnston, A. W. B. & Downie, J. A. *EMBO J.* **9**, 349-354 (1990).
- Long, S. R. & Atkinson, E. M. *Nature* **344**, 712-713 (1990).
- Truchet, G., Michel, M. & Denarie, J. *Differentiation* **16**, 163-172 (1980).
- Hirsch, A. M., Bhuvaneshwari, T. V., Torrey, J. G. & Bisseling, T. *Proc. natn. Acad. Sci. U.S.A.* **86**, 1244-1248 (1989).
- Long, S. R. & Cooper, J. in *Molecular Genetics of Plant-Microbe Interactions* (eds Verma, D. P. S. & Palacios, R.) 163-178 (American Phytopathological Society, St. Paul, MN, 1988).

72. Dehio, C. & de Bruijn, F. J. *Pl. J.* **2**, 117–128 (1991).
73. Koblika, B. K. *et al.* *Science* **240**, 1310–1316 (1988).
74. Buck, L. & Axel, R. *Cell* **65**, 175–187 (1991).
75. Vance, C. P. *Rev. Microbiol.* **37**, 399–424 (1983).
76. Vance, C. P., Johnson, L. E. B. & Hardarson, G. *Physiol. Pl. Pathol.* **17**, 167–173 (1986).
77. Djordjevic, M. A., Gabriel, D. W. & Rolfe, B. G. A. *Rev. Phytopath.* **25**, 145–168 (1987).
78. Parniske, M., Fischer, H.-M., Hennecke, H. & Werner, D. *Z. Naturforsch.* **46**, 318–320 (1991).
79. Keen, N. T. & Staskawicz, B. A. *Rev. Microbiol.* **42**, 421–440 (1988).
80. Morris, R. O. A. *Rev. Pl. Physiol.* **37**, 509–538 (1986).
81. Shen, W. H., Petit, A., Guern, J. & Tempe, J. *Proc. natn. Acad. Sci. U.S.A.* **85**, 3417–3421 (1988).
82. Truchet, G. *et al.* *Molec. Gen. Genet.* **219**, 65–68 (1989).
83. Ehrhardt, D. W., Atkinson, E. M. & Long, S. R. *Science* **256**, 998–1000 (1992).
84. Innes, R. W. *et al.* *Molec. Gen. Genet.* **201**, 426–432 (1985).
85. Göttfert, M., Weber, J. & Hennecke, H. *J. Pl. Physiol.* **132**, 394–397 (1988).
86. Goethals, K., Gao, M., Tomekpe, K., Van Montagu, M. & Holsters, M. *Molec. Gen. Genet.* **219**, 289–298 (1989).
87. Hong, G.-F., Burr, J. E. & Johnston, A. W. *Nucleic Acids Res.* **15**, 9677–9690 (1987).
88. Fisher, R. F., Egelhoff, T. T., Mulligan, J. T. & Long, S. R. *Genes Dev.* **2**, 282–293 (1988).
89. Downie, J. A. *Molec. Microbiol.* **3**, 1649–1651 (1989).
90. Surin, B. P. & Downie, J. A. *Molec. Microbiol.* **2**, 173–183 (1988).
91. Canter Cremers, H. C. J. *et al.* *Pl. molec. Biol.* **13**, 163–174 (1989).
92. Göttfert, M., Grob, P. & Hennecke, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2680–2684 (1990).
93. Davis, E. O., Evans, I. J. & Johnston, A. W. B. *Molec. Gen. Genet.* **212**, 531–535 (1988).
94. Scott, K. F. *Nucleic Acids Res.* **14**, 2905–2919 (1986).
95. Nieuwkoop, A. J. *et al.* *J. Bact.* **169**, 2631–2638 (1987).
96. Stacey, G. *et al.* in *Advances in Molecular Genetics of Plant–Microbe Interactions* (eds Hennecke, H. & Verma, D. P. S.) 156–161 (Kluwer, Dordrecht, 1991).
97. Sadowsky, M. J. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 637–641 (1990).
98. Davis, E. O. & Johnston, A. W. B. *Molec. Microbiol.* **4**, 921–932 (1990).
99. Swanson, J. A., Mulligan, J. & Long, S. R. *Genetics* (in the press).
100. Bibb, M. J., Biro, S., Motamedi, H., Collins, J. F. & Hutchinson, C. R. *EMBO J.* **8**, 2727–2738 (1989).
101. Kondorosi, E. *et al.* *Molec. Microbiol.* **5**, 3035–3048 (1991).

ACKNOWLEDGEMENTS. We wish to thank all our colleagues who generously responded to our request for recent reprints and communication of results before publication. Work from our laboratory was funded by the NIH and the Department of Energy.

ARTICLES

Anisotropy measurements of the cosmic microwave background radiation at intermediate angular scales

R. A. Watson*, C. M. Gutierrez de la Cruz†, R. D. Davies*, A. N. Lasenby‡
R. Rebolo†, J. E. Beckman† & S. Hancock‡

* University of Manchester, Nuffield Radio Astronomy Laboratories, Jodrell Bank, Macclesfield SK11 9D1, UK

† Instituto de Astrofísica de Canarias, 38071 La Laguna, Tenerife, Spain

‡ Mullard Radio Astronomy Observatory, Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK

Galaxy formation by gravitational instability implies the presence of temperature fluctuations in the cosmic microwave background. New observations at 10.45 and 14.9 GHz from the Observatorio del Teide in Tenerife on angular scales of several degrees reveal structure of galactic origin at the lower frequency, while at the upper frequency the magnitude of fluctuations of cosmological origin must be less than 1.8×10^{-5} on a scale of 5° . This strongly constrains models of galaxy formation, and is compatible with the recently announced discovery by the COBE satellite of fluctuations on larger angular scales.

MANY competing theories have been proposed to describe the formation of structure in the Universe. Observations of the cosmic microwave background (CMB) can provide a direct and unique indication of the amplitudes and scale-sizes of the primordial density perturbations that eventually grow to become the structure in the present Universe. Current theories are coming under increasing pressure from CMB observations and from studies of large-scale structures and bulk flows in the distribution of galaxies. Some of the strongest constraints are provided by CMB studies of perturbations on angular scales larger than the horizon scale at recombination ($\geq 2^\circ$) which unambiguously probe the long-wavelength portion of the intrinsic perturbation spectrum. The angular scales associated with the precursors of features actually known to exist now are much smaller, ranging from subarcminute scales for galaxies to 5–10 arcminutes for a rich cluster of galaxies and a few tens of arcminutes for super-clusters. Even the very-large-scale structures recently found in galaxy clustering and redshift surveys^{1–4} correspond to angular scales of only 0.5 – 1° . Thus expectations of structure in the CMB on larger scales are based on theories for the physical mechanism that produces the seed fluctuations in the early Universe. If this mechanism is the amplification of quantum irregularities by inflation, for example, then a ‘Harrison–Zeldovich’ or scale-invariant spectrum of density perturbations is thought to be the likely result. The effects of a particular perturbation spectrum

on the CMB are best expressed through the angular autocovariance function or its power spectrum equivalent. The former is defined by

$$C(\theta) = \left(\frac{\Delta T}{T}(n_1) \frac{\Delta T}{T}(n_2) \right)$$

where n_1 and n_2 are two directions on the sky separated by angle θ , and the average is taken over the whole sky. Specific forms for $C(\theta)$ will be assumed in the analysis below. The power spectrum can be expressed in terms of statistical spherical harmonic coefficients a_l^2 defined by

$$C(\theta) = \frac{1}{4\pi} \sum_{l \geq 2}^{\infty} a_l^2 (2l+1) P_l(\cos \theta)$$

where for a Harrison–Zeldovich spectrum $a_l^2 \propto 1/l(l+1)$. The amplitude of the fluctuations is predicted by comparisons with present-day galaxy clustering, and therefore in models with a dark matter component these predictions rest on an assumption about the degree to which the distribution of light fairly traces the density perturbations of the underlying dark matter. This is expressed by a bias parameter b , and roughly speaking, predicted CMB fluctuations scale as b^{-1} . The amplitude depends also on the mode of the perturbations in the early Universe. These can be either adiabatic, with matter and radiation perturbations

linked by $\Delta T/T = \frac{1}{3}\Delta\rho/\rho$, or isocurvature, for which the total energy density is constant. The latter are expected to give enhanced CMB anisotropies on large angular scales. Our new Tenerife experiments⁵ provide high-sensitivity data covering a substantial area of the sky (~ 1 sr) at angular harmonics $12 \leq l \leq 30$ (see Fig. 1). These results complement those from the differential microwave radiometer equipment on the COBE satellite⁶, which provides all-sky coverage at lower sensitivity per beam over angular harmonics $1 \leq l \leq 15$.

The most popular scenario of galaxy formation is the cold dark matter (CDM) theory; but this has had difficulty in explaining the presence of very-large-scale structure in the nearby Universe¹⁻⁴. Furthermore, attempts to overcome this problem by putting additional large-scale power into the initial perturbation spectrum⁷ encounter difficulties with the current limits on the anisotropy of the CMB^{5,6,8-12} on intermediate and large scales. A further modest reduction in these experimental limits would put pressure even on unmodified CDM theories^{10,11}.

To achieve these more sensitive levels, the noncosmological backgrounds such as diffuse galactic emission^{13,14} and extragalactic radio source confusion¹⁵ must be measured and accounted for. These components can be separated from the CMB perturbations by multifrequency observations. We demonstrate a substantial improvement in the sensitivity to CMB perturbations by measurements with identical beamwidths at the two frequencies 10.45 and 14.9 GHz. Our original observations⁵ at 10.45 GHz detected structure at a level of $\Delta T/T = 3.7 \times 10^{-5}$ with an 8.4° beam surveying an area of declination $\delta = 40^\circ$ between right ascension $180^\circ < \alpha < 250^\circ$. The main signal occurred at $\alpha = 220^\circ$; the presence of this 'hotspot' was confirmed by independent analysis^{16,17} of our data. Our new observations with a resolution of 5.6° confirm the earlier detection and show that it is of galactic origin. We are then able to use the 14.9-GHz data to set very low limits to the CMB perturbation amplitude.

Observations and data reduction

An important consideration in the design of a CMB anisotropy experiment is the sensitivity of the beam configuration to structure on various angular scales. Figure 1 shows the filter function of our original 8.4° beamwidth (full width at half power, FWHP) 10.45-GHz experiment in comparison with those of the COBE satellite⁶ and South Pole experiment⁹. Our original 10.45-GHz experiment was modified to a 5.6° beamwidth, retaining the 8.1° beamswitch. The modification to a smaller beamwidth improves the sensitivity and, equally important, gives a closer match to the predictions of CDM models. The 14.9-GHz experiment has the same beam configuration as the modified 10.45-GHz experiment.

We obtained a triple-beam configuration⁵ by fast-switching at 63 Hz between two horns, combined with a rocking plate which threw the double beam a distance of $\pm 8.1^\circ$. The resultant configuration consisted of a positive beam located on the meridian and two half-amplitude negative beams at declinations close to that of the central beam. The rocking plate directed the beam pairs (primary differencing) for a period of 4 s to the east followed by 4 s to the west to provide secondary differencing; every 82 s, data were recorded consisting of 8 pairs of secondary differences. Scans were made using Earth rotation at fixed elevation to obtain repeated coverage of 24 h in right ascension at fixed declination. A system of continuous on-line calibration using noise diodes was supplemented by daily astronomical calibration on the galactic plane crossings and, every few months, by Moon scans to establish an absolute scale. The system parameters at 10.45 and 14.9 GHz were $T_{\text{sys}} = 85$ K, $\Delta\nu = 470$ MHz and $T_{\text{sys}} = 75$ K, $\Delta\nu = 900$ MHz, respectively. These give sensitivities of 8 and 5 mK Hz^{-1/2} respectively after allowance for a factor of 2 lost in the double switching scheme. Because each system had two independent channels, a factor of $\sqrt{2}$ in sensitivity was recovered. For this work we used the

proven good site at Observatorio del Teide on Tenerife at an altitude of 2,390 m.

The first step in data analysis was to remove data when the Sun or Moon were closer than 50° to the centre of the main beam, or when the noise on the scan produced by cloud or high humidity was increased by a factor of more than 2. The next step was to eliminate long-period baseline drifts on angular scales greater than those to which the beam configuration was sensitive¹⁸; such effects were more pronounced in the 14.9-GHz data, chiefly because of greater atmospheric emission. To subtract the baseline, we simultaneously processed all scans by a maximum entropy method (MEM) algorithm, which could separate drifts from real sky signal by the creation of an estimated sky emission model. For 10.45 GHz, data were available from adjacent declinations separated by half a beamwidth. This allowed greater discrimination by the model and the production of a two-dimensional map of the microwave background emission.

The final scan at $\delta = 40^\circ$ for each beam configuration was produced by the weighted average of the individual scans with the baseline subtracted; the weighting of each scan was inversely proportional to the variance of the scan. The high-latitude data at 10.45 GHz and 14.9 GHz with a resolution of 5.6° are shown in Fig. 2 for the region $155^\circ \leq \alpha \leq 250^\circ$. The average root mean square (r.m.s.) value per 1° of right ascension α was 270 μ K at 10.45 GHz and 96 μ K at 14.9 GHz; the corresponding standard errors per beam are 97 μ K and 25 μ K respectively. The greater sensitivity of the 14.9-GHz experiment is due to its greater bandwidth and number of scans.

Radio-source and galactic predictions

The new data for 5.6° beamwidth and 10.45 GHz plotted in 1° intervals in Fig. 2 show evidence for emission peaks centred near $\alpha = 185^\circ$, 225° and 250° . We now compare this structure from our three data sets with predictions of discrete radio sources and galactic emission for the high latitude ($155^\circ < \alpha < 260^\circ$) region of the $\delta = 40^\circ$ scan. Column 1 of Fig. 3 shows the predictions of the discrete source and galactic emission components separately for the scans at 10.45 GHz and 8.4° , 10.45 GHz and 5.6° , and 14.9 GHz and 5.6° . The discrete source data are taken from ref. 19; for a given beamwidth the observed temperature $T_b \propto \nu^{-2-\alpha}$ where α is the flux spectral index (flux $S \propto \nu^{-\alpha}$). The only significant contributor is the flat-spectrum variable

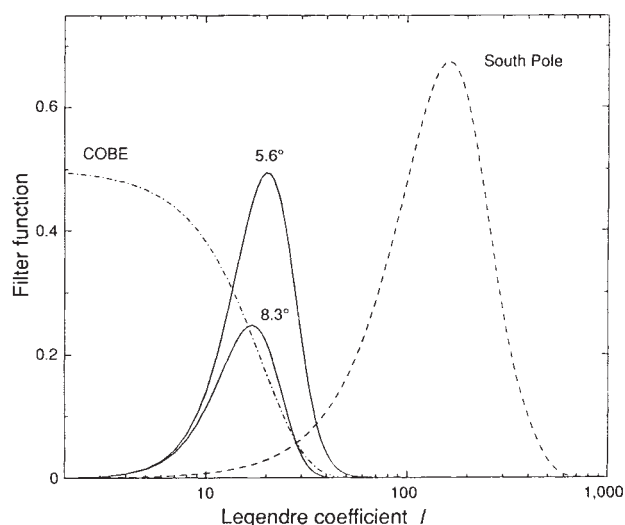


FIG. 1 Plot of relative sensitivities of the beam configurations of our Tenerife 8.4° and 5.6° experiments compared with the COBE and South Pole systems. The sensitivity is calculated in terms of the Efstathiou filter function against Legendre coefficient l .

source 3C345 at $\alpha = 250^\circ$ which undoubtedly accounts for the peaks in the three data sets at this point. This agreement confirms that our experimental technique and analysis procedures are basically sound. The galactic emission is derived from published 408-MHz and 1,420-MHz data (refs 20 and 21, respectively) using the locally derived spectral index (in the range 2.6 to 2.8; ref. 22).

To make comparisons with galactic extrapolations and point-source catalogues it is necessary to convolve all of them, including the MEM-produced sky emission map, with the triple-beam pattern. The resulting continuous curves for the three data sets are shown in column 2 of Fig. 3; superposed are the sum of the discrete-source and galactic emission predictions from column 1. The three data sets are remarkably consistent with each other. The two 10.45-GHz scans show similar amplitude of structure, although the 5.6° data are larger as would be expected from the larger horn aperture. The 14.9-GHz structure, although it agrees in general shape, is more than a factor of 3 smaller, corresponding to a spectral index $\alpha \approx 3.1$. Such high-frequency steepening is expected for galactic synchrotron emission^{13,23–25}.

A feature of column 2 in Fig. 3 is the poor agreement between the observations and predictions from low-frequency data. The only common feature is that at $\alpha = 225^\circ$; it is however a factor of 2 smaller than predicted, as expected for galactic spectral steepening. It is clear that the extrapolation of spectral indices between 408 and 1,420 MHz is not valid for frequencies of 10 GHz and above. Some of the uncertainty occurring in the predictions actually arises from baseline artefacts in the low-frequency data (R.D.D. *et al.*, manuscript in preparation). There is clearly no substitute for sensitive multifrequency observations at these higher frequencies if CMB anisotropies are to be investigated and quantified.

Our conclusion is that galactic emission can be identified on scales of 5–8° with the sensitivity of our present 10.45-GHz observations. The spectral index at this frequency seems to be steeper than 3.0 between 10.45 and 14.9 GHz; the further improvement in sensitivity which is possible with the present 14.9-GHz experiment (a factor of 2–3) should define this index more clearly. In principle it would be possible to correct the 14.9-GHz data for galactic emission by assuming a spectral

index (of say 3.0) and then quantify the CMB perturbation amplitude. We have taken a more conservative approach in not removing any galactic contributions; even so, it is possible to set a sensitive upper limit to CMB perturbations.

Analysis of fluctuation amplitudes

The 5.6° beamwidth data at 10.45 and 14.9 GHz were analysed using the method of likelihood quotients followed by bayesian integration with a uniform prior^{10,18,26,27}; other types of analysis give similar results. The intrinsic sky fluctuations are characterized by a random gaussian field described by a one-dimensional autocorrelation function similar to that used in refs 5 and 27, namely $C(\theta) = C_0 \exp(-\theta^2/(2\theta_c^2))$, which after convolution with the radiotelescope beam is transformed to

$$C_M(\theta, \theta_c, \sigma) = C_0 \frac{\theta_c^2}{(2\sigma^2 + \theta_c^2)} \exp\left(\frac{-\theta^2}{2(2\sigma^2 + \theta_c^2)}\right)$$

Here θ_c is the coherence angle on the sky, and σ is the dispersion of the beam ($\sigma = 2.4^\circ$ for the beam of 5.6° FWHP). The likelihood function is

$$L \propto \frac{1}{\sqrt{|V|}} \exp\left(-\frac{1}{2} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)$$

where $\mathbf{X}$ is the vector formed by the second differences (of the triple-beam response) and $\mathbf{V}$ the covariance matrix, which itself is the sum of two covariance matrices, the instrumental and the trial covariance expected in the data (see for example equation

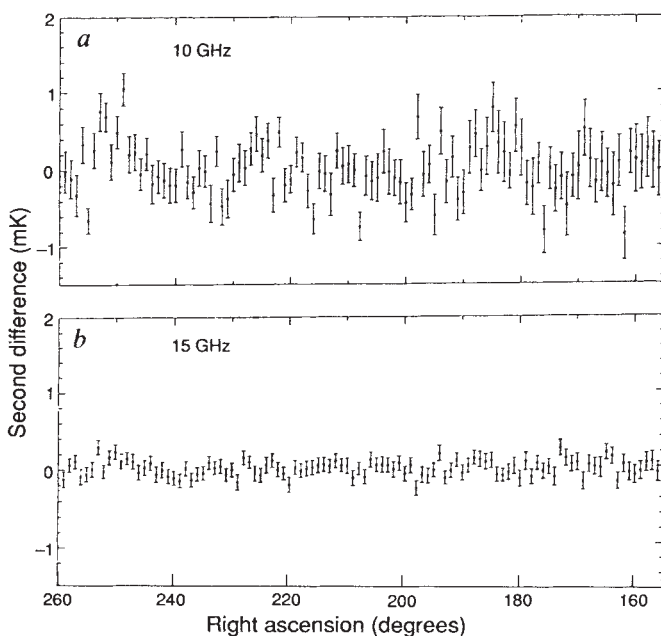


FIG. 2 *a*, 10.45- and *b*, 14.9-GHz co-added scans with a beamwidth of 5.6° at $\delta = 40^\circ$ for the right ascension range $155^\circ \leq \alpha \leq 250^\circ$. The double-switching (triple beam) data are plotted every 4 min (1°) in right ascension.

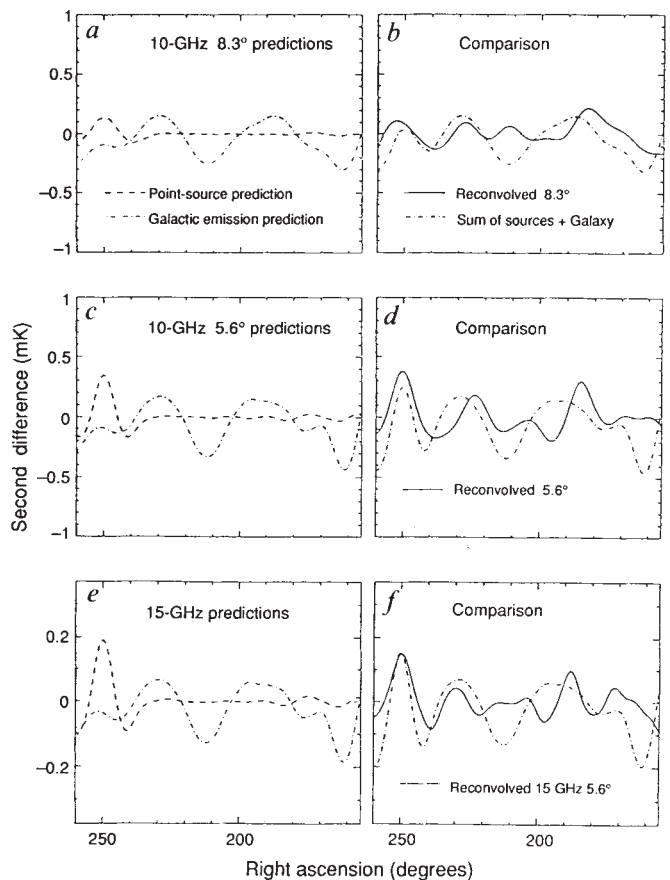


FIG. 3 The galactic and discrete source predictions (left-hand plots) compared with the observations (right-hand plots) for the 10.45-GHz 8.4° and 5.6° data and for the 14.9-GHz data assuming a brightness temperature spectral index of 2.8. The discrete sources are clearly seen in the data whereas the 408-MHz galactic prediction is only poorly represented by the data.

(6) of ref. 16). This function is then plotted against the two parameters C_0 and θ_c to give evidence for detection which will manifest itself as a peak away from the zero axes in the (C_0, θ_c) plane.

Figure 4 shows cuts through the $L(C_0, \theta_c)$ surface at a fixed θ_c for the data at 5.6° beamwidth, and frequencies 10.45 and 14.9 GHz, plotted in terms of $\sqrt{C_0}$ which is the intrinsic r.m.s. of the sky (in μK). The plot for 10.45 GHz (Fig. 5a) is made at a coherence angle $\theta_c = 3.4^\circ$ where the likelihood maximizes and covers the source-free region $\alpha = 161^\circ$ to 234° , avoiding the radio sources at $\alpha = 250^\circ$. The 14.9-GHz plot (Fig. 5b) was made for $\theta_c = 4^\circ$, where the observing system has maximum sensitivity, and covers the region $\alpha = 161$ – 260° ; we corrected for the sources at $\alpha = 250^\circ$ by using the published^{19,28} 15-GHz fluxes.

At 10.45 GHz we have a clear detection with a likelihood ratio (LR) of 10.3 at $\sqrt{C_0} = 150 \mu\text{K}$ which corresponds to the intrinsic fluctuation $(\Delta T/T)_{\text{intrinsic}} = 5.7 \times 10^{-5}$ that we reported for our 8° beamwidth data⁵, again confirming the consistency of our results. The corresponding r.m.s. smoothed in a gaussian beam of FWHP = 5.6° is $\sqrt{C_M(\theta=0)} = 106 \mu\text{K}$ or $(\Delta T/T)_{5.6^\circ \text{ beam}} = 3.9 \times 10^{-5}$. By contrast, the 14.9-GHz plot shows no detection, with the LR falling smoothly to zero from unity at $C_0 = 0$. A bayesian integration under the LR curve, using a uniform prior, set an upper limit at 95% confidence of $\sqrt{C_0} < 51 \mu\text{K}$, corresponding to $(\Delta T/T)_{\text{intrinsic}} < 1.8 \times 10^{-5}$. The choice of uniform prior sets a conservative upper limit (see ref. 18), and the use of intrinsic r.m.s. is itself conservative compared with the common practice of quoting results for $\Delta T/T$ after smoothing by the beam (see for example ref. 8). Our limit, one of the lowest reported for any CMB fluctuation search, can be used to set new constraints on theories of the origin of structure in the early Universe.

Comparison with theory

To compare our experimental upper limit for CMB fluctuations with theoretical predictions in detail, we need to calculate the autocovariance function for the galaxy formation scenario to be used in the likelihood and bayesian analysis of the data; this full discussion will be given elsewhere (A.N.L., manuscript in preparation). Here we outline the effect that our upper limit, in combination with recent data on other angular scales, has on the various categories of theory that have survived previous CMB constraints. Models of galaxy formation that involved universes containing only baryonic matter have been ruled out by previous observations for any density parameter Ω ; the new data exclude them even more strongly.

Cold dark matter models. In Fig. 5 we compare our 14.9-GHz upper limit to the CMB anisotropy with those from other recent experiments. The plot also shows the anisotropy expected from a CDM model, taking proper account of the beam geometry of each experiment. The predictions are based on the autocovariance function corresponding to the standard model²⁹ with $\Omega = 1$, $n = 1$ (the Harrison-Zeldovich spectrum), $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and $b = 1.8$ (standard CDM biasing). The gaussian and the Harrison-Zeldovich spectra give similar results, as found previously¹¹. In general, the predicted $\Delta T/T$ peaks at an angular scale near 1° , reflecting the increased power in perturbations on these scales in standard CDM models. The predicted values of CDM $\Delta T/T$ in the Tenerife 5.6° and the South Pole 1° experiments are 6×10^{-6} and 10×10^{-6} , respectively, compared with the observed limits of 1.8×10^{-5} and 3.5×10^{-5} . If we measure sensitivity as the ratio of the observed upper limit to the CDM prediction, our own result is rather more constraining than the South Pole result. We can use our result to set a constraint on the biasing parameter b , as the amplitude of the CMB anisotropies varies inversely as b . Vittorio *et al.*¹¹ find that $\Omega = 1$ CDM with no biasing ($b = 1.0$) predicts a value 1.88 below the South Pole limit, whereas our result for $b = 1.8$ is a factor of 3 above the predictions. We thus deduce a lower limit for $b > 0.6$. This may be compared with the upper

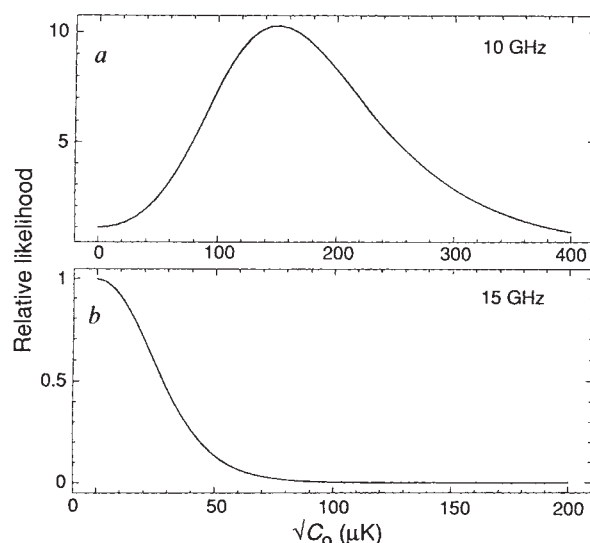


FIG. 4 Likelihood curves for (a) the 10.45- and (b) the 14.9-GHz 5.6° data at Dec = 40° in the right ascension range $160^\circ < \alpha < 250^\circ$. $\sqrt{C_0}$ gives the intrinsic r.m.s. level of the sky emission. A significant signal is seen in the 10.45-GHz data, whereas no significant detection above the noise is seen in the 14.9-GHz data. Note the different $\sqrt{C_0}$ scales in the two plots.

limit set from observations of large-scale bulk flows^{30,31}, at a level of $b < 1.4$ at 1.4 at 95% confidence, from the analysis of Heavens³². An increase in sensitivity of either the Tenerife or South Pole observations by a factor of ~ 2 will put strong pressure on CDM models with any bias.

To decide between nonstandard CDM models (those which depart from the inflationary values of $\Omega = 1$ and $n = 1$, or attempt to modify the CDM spectrum) the compilation of Suto *et al.*³³ is useful. The present 14.9-GHz result from Tenerife is 3.0 times more constraining than the value taken by Suto *et al.* from the 10.45-GHz result⁵. Previously, our large-angular-scale data and that of ref. 8 required $n > 0$, regardless of the value of Ω . The new limits imply $n > 0.7$, that is less power on large scales. But increased power is in fact observed on large scales in the APM¹ and QDOT⁴ surveys, where structure in the 30–50 Mpc range can be expected to feed through to enhanced CMB anisotropies on scales of $\sim 1^\circ$. This does not necessarily mean that anisotropies will be enhanced on the 5° scale of our experiment, especially if artificial cut-offs have been imposed to avoid this (see ref. 34). Our comments above have referred to CDM with adiabatic fluctuations. If isocurvature perturbations are used instead, then CDM models were already virtually ruled out³³ by large-scale observations. The reduction by a factor of 3 makes it even more certain that these models are not viable.

Baryonic isocurvature with reionization. Baryonic isocurvature universes with early reionization³⁵ have been considered in the context of CMB anisotropies^{36–40}. The discussion centres on the power spectrum slope n and the density Ω allowed by the CMB constraints. Gouda *et al.*³⁸ show that the results on intermediate angular scales rule out all $n < -1$ in the notation of Efstathiou and Bond³⁶. This of course rules out the inflationary value of $n = -3$, although galaxy clustering observations suggest³⁹ $n \approx -1$. By using the approach of Efstathiou³⁷ we show the constraints that our new 14.9-GHz results place within the (Ω, n) plane in Fig. 6. The previously allowable region with $n \approx -1$, $\Omega = 1$ is now virtually rejected and the maximum Ω allowed is ~ 0.7 , making Peebles models³⁵ with zero cosmological constant very difficult to reconcile with inflation.

Non-zero cosmological constant. The constraints provided by CMB anisotropy limits on universes with non-zero cosmological constant Λ have been surveyed by Sujiyama *et al.*³⁹. In the context of inflation, it is natural to require that Λ when combined with Ω should give a model with zero spatial curvature

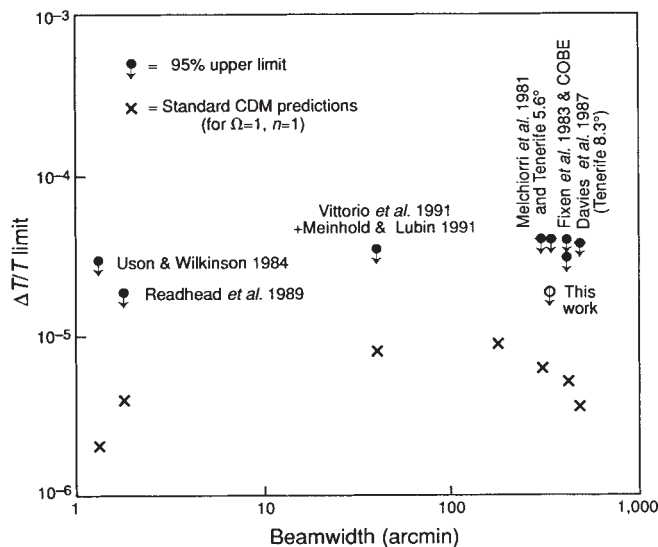


FIG. 5 Observed limits on $\Delta T/T$ as a function of beamwidth (full width half maximum) for the most sensitive searches for CMB fluctuations. The predictions for each observation on the standard CDM model ($\Omega=1$, $n=1$, $h=0.5$, $b=1.8$) are plotted as crosses.

($\Lambda + \Omega = 1$). Using Ω as the single free parameter and assuming the inflationary spectrum $n=1$, Sugiyama *et al.* consider in turn baryonic adiabatic and isocurvature models, hot dark matter (HDM) and CDM adiabatic and isocurvature models, and, in addition, the Peebles model (further discussed in ref. 40). We can now apply to their analysis the new 14.9-GHz result for the large-scale limit and adopt, as they did, the result of Readhead *et al.*²⁷ for the limit at small angular scales. All the isocurvature models with a non-zero cosmological constant are ruled out by our new limit, as are all baryonic adiabatic models at all acceptable values of the Hubble constant H_0 . Three possible cases with non-zero Λ may remain. The first is CDM with adiabatic perturbations. The second is HDM with adiabatic perturbations, for which Sugiyama *et al.* give no predictions on our scale sizes; Anninos *et al.*⁴¹ suggest values of $\Delta T/T = 7\text{--}14 \times 10^{-6}$, only just below our observed limit. The third is the Peebles model with $n > 1$ for which Sugiyama *et al.* do not give predictions. Yokoyama and Suto⁴⁰ address the third scenario; their predictions are in conflict with the present upper limit when combined with the upper limits of Readhead *et al.* The elimination of all the isocurvature models with a cosmological constant is a relief in some respects, as they all involved unrealistically short ages for the Universe (< 7.8 Gyr for HDM with isocurvature, or < 9 Gyr for baryonic or CDM isocurvature). Our result shows the potential difficulties in schemes⁴² that attempt to reconcile large-scale galaxy clustering with CDM by including a cosmological constant.

Constraints on strings, lumps and voids

Measurements of the CMB anisotropy can in principle also be used to constrain the cosmic-string model for galaxy formation^{43–46}. The constraints are of the same order as limits obtained by gravitational radiation⁴⁶, and are in the range that would be required if cosmic strings were the principal nucleating agents for galaxy formation. An advantage of our mode of observation, which covers a large continuous range in right ascension at neighbouring declinations, is an enhanced sensitivity to non-gaussian fluctuations of the kind that strings will generate. This makes it potentially useful for detecting the imprints left by individual density inhomogeneities; evidence of this has been accumulating for density inhomogeneities on large scales in the Universe⁴⁷ (for example the Bootes void, Great Attractor, Shapley concentration, and Great Wall). The

gravitational effect of such structures on the background radiation has been studied^{48–51}. In general their effect is to induce temperature fluctuations of order $\Delta T/T \approx 10^{-6}$ or smaller. But one can envisage cases⁵¹, such as a compensated void of radius $100 h^{-1}$ Mpc at a redshift $z \leq 1$, or galaxy concentrations of $\sim 10^{17}$ solar masses, for which the effects could reach the sensitivity of our experiment ($\sim 10^{-5}$). Although more integration time or a gain in sensitivity is needed to prove these effects, our current upper limit already imposes serious constraints on the existence of nonlinear structures.

Conclusion

Our CMB measurements enable us to identify the emission from the Galaxy and assign upper limits to the CMB component with high confidence. The galactic emission structure at high latitudes has a spectral index $\alpha \approx 3.1$. Ideally, a third frequency is required to define the galactic spectral index and brightness at each point surveyed; we are already engaged in observations at 33 GHz with the same angular resolution as the 14.9-GHz experiment and twice the sensitivity. We expect to map features in the cosmic background with a brightness $\Delta T/T = 5\text{--}7 \times 10^{-6}$ at 95% confidence. Observations at 10.45 and 14.9 GHz are continuing, and with a more extensive data set, we expect to improve on the 14.9-GHz sensitivity by a factor of 2–3.

The limit given here on CMB anisotropy ($\Delta T/T_{\text{intrinsic}} < 1.8 \times 10^{-5}$) is the most sensitive on scales of a few degrees, and the most restrictive on any scale. When combined with the smaller angular scale measurements of Readhead *et al.*²⁷, it eliminates several of the commonly considered scenarios of galaxy formation, leaving as possibilities CDM scenarios with $n > 0.7$; even these models have difficulty in accounting for large-scale structure and streaming. The sensitivity of CMB fluctuation experiments is reaching a stage where new cosmological thinking may be required.

Since the submission of this manuscript, the analysis of the first year of observation by the Cosmic Background Explorer (COBE) has been announced. A detection of $30 \mu\text{K}$, corresponding to $(\Delta T/T)_{\text{measured}} = 1.1 \times 10^{-5}$, has been claimed on a 10° scale on the basis of an autocorrelation of 2,200 data points on a 2.6° grid. This value is a factor of 1.5 below our upper limit of $51 \mu\text{K}$ (at the 95% confidence level) on a 5° scale. The COBE result is statistically consistent with CDM fluctuations having a low bias value ($b \approx 1$) and a Harrison-Zeldovich spectrum ($n=1$), thus raising all the CDM predictions shown in Fig. 5 by a factor of 1.8. If the COBE result is confirmed and the

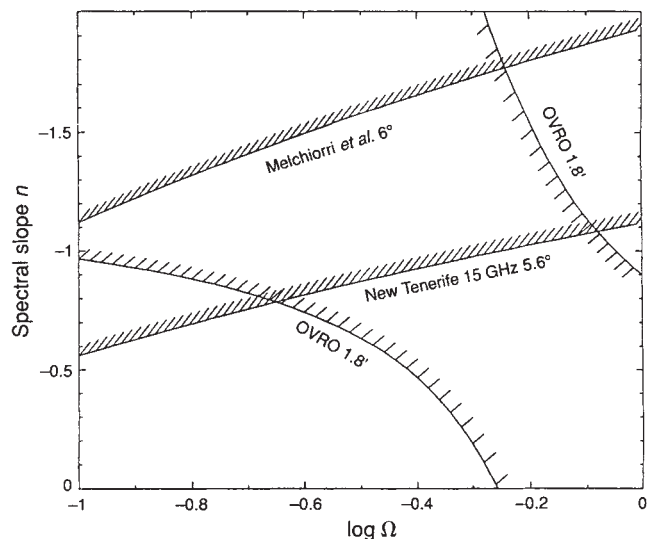


FIG. 6 Constraints in the Ω , n plane set by the new 14.9-GHz data in combination with those of Readhead *et al.*²⁷ (Owens Valley Radio Observatory) and Melchiorri *et al.*⁸. Hatching shows the areas excluded by the observations for a re-ionized Universe³⁵.

spectrum is Harrison–Zeldovich, our current sensitivity of 35 μK (standard error) in a single 5° beam is 1.5 times greater than the expected three-beam fluctuation level on this scale ($\sim 24 \mu\text{K}$). The corresponding sensitivity of COBE in its 7° beam

is 75 μK . Detection of actual features in the CMB structure remains the main aim of our Tenerife work. In the next year, we expect to make the factor of 2–3 improvement in sensitivity and to delineate actual structural features. $\square$

Received 9 March; accepted 19 May 1992.

1. Maddox, S. J., Efsthathiou, G., Sutherland, W. J. & Loveday, J. *Mon. Not. R. astr. Soc.* **242**, 43p–47p (1990).
2. Geller, M. J. & Huchra, J. P. *Science* **246**, 897–903 (1990).
3. Broadhurst, T. J., Ellis, R. S., Koo, D. C. & Szalay, A. S. *Nature* **343**, 726–728 (1990).
4. Saunders, W. *et al.* *Nature* **349**, 32–38 (1991).
5. Davies, R. D. *et al.* *Nature* **326**, 462–465 (1987).
6. Smoot, G. F. *et al.* *Astrophys. J. Lett.* **371**, L1–L5 (1991).
7. Salopek, D. S., Bond, J. R. & Bardeen, J. M. *Phys. Rev. D* **40**, 1753–1788 (1989).
8. Melchiorri, F., Melchiorri, B. O., Ceccarelli, C. & Pietranera, L. *Astrophys. J. Lett.* **250**, L1–L4 (1981).
9. Meinhold, P. & Lubin, P. *Astrophys. J. Lett.* **370**, L11–L14 (1991).
10. Bond, J. R., Efsthathiou, G. P., Lubin, P. M. & Meinhold, P. R. *Phys. Rev. Lett.* **66**, 2179–2182 (1991).
11. Vittorio, N., Meinhold, P., Muciaccia, P. F., Lubin, P. & Silk, J. *Astrophys. J. Lett.* **372**, L1–L4 (1991).
12. Meyer, S. S., Cheng, E. S. & Page, L. A. *Astrophys. J. Lett.* **371**, L7–L9 (1991).
13. Banday, A. J. & Wolfendale, A. W. *Mon. Not. R. astr. Soc.* **245**, 182–191 (1990).
14. Masi, S. *et al.* *Astrophys. J. Lett.* **366**, L51–L55 (1991).
15. Franceschini, A., Toffolatti, L., Danese, L. & De Zotti, G. *Astrophys. J.* **344**, 35–45 (1989).
16. Vittorio, N., de Bernardis, P., Masi, S. & Scaramella, R. *Astrophys. J.* **341**, 163–167 (1989).
17. Vittorio, N. in *Large-Scale Motions in the Universe* (eds Rubin, V. C. & Coyne, G. V.) 397–417 (Vatican Press and Princeton, 1988).
18. Lasenby, A. N. & Davies, R. D. in *Large-scale Motions in the Universe* (eds Rubin, V. C. & Coyne, G. V.) 277–297 (Vatican Press and Princeton, 1988).
19. Kühr, H., Witzel, A., Pauliny-Toth, I. I. K. & Nauber, U. *Astr. Astrophys. Suppl.* **45**, 367–430 (1981).
20. Haslam, C. G. T., Salter, C. J., Stoffel, H. & Wilson, W. E. *Astr. Astrophys. Suppl.* **47**, 1–143 (1982).
21. Reich, P. & Reich, W. *Astr. Astrophys. Suppl.* **74**, 7–28 (1988).
22. Lawton, K. D., Mayer, C. J., Osborne, J. L. & Parkinson, M. L. *Mon. Not. R. astr. Soc.* **225**, 307–327 (1987).
23. Banday, A. J. & Wolfendale, A. W. *Mon. Not. R. astr. Soc.* **248**, 705–714 (1991).
24. Watson, R. A., Rebol, R., Beckman, J. E., Davies, R. D. & Lasenby, A. N. in *Large-scale Structures and Motion in the Universe* (eds Mezzetti, M., Giuricin, G., Mardirossian, F. & Ramella, M.) (Kluwer, Dordrecht, 1989).

25. Banday, A. J., Giler, M., Szabelska, B., Szabelski, J. & Wolfendale, A. W. *Astrophys. J.* **375**, 432–438 (1991).
26. Lawrence, C. R., Readhead, A. C. S. & Myers, S. T. in *The Post-Recombination Universe* (eds Kaiser, N. & Lasenby, A. N.) 173–181 (Kluwer, Dordrecht, 1988).
27. Readhead, A. C. S. *et al.* *Astrophys. J.* **346**, 566–587 (1989).
28. Gregory, P. C. & Condon, J. J. in *The 87GB Catalog of Radio Sources at 4.85 GHz* (NRAO/AUI, Charlottesville, 1990).
29. Bond, J. R. & Efsthathiou, G. *Mon. Not. R. astr. Soc.* **226**, L66–L87 (1987).
30. Lynden-Bell, D. *et al.* *Astrophys. J.* **326**, 19–49 (1988).
31. Staveley-Smith, L. & Davies, R. D. *Mon. Not. R. astr. Soc.* **241**, 787–826 (1989).
32. Heavens, A. F. *Mon. Not. R. astr. Soc.* **251**, 267–280 (1991).
33. Suto, X., Gouda, N. & Sugiyama, N. *Astrophys. J. Suppl.* **74**, 665–674 (1990).
34. Efsthathiou, G. in *Observational Tests of Cosmological Inflation* (eds Shanks, T., Banday, A. J., Ellis, R. S., Frenk, C. S. & Wolfendale, A. W.) 425–436 (Kluwer, Dordrecht, 1991).
35. Peebles, P. J. E. *Astrophys. J. Lett.* **315**, L73–L76 (1987).
36. Efsthathiou, G. & Bond, J. R. *Mon. Not. R. astr. Soc.* **227**, 33p–38p (1987).
37. Efsthathiou, G. in *Large-Scale Motions in the Universe* (eds Rubin, V. C. & Coyne, G. V.) 299–319 (Princeton Univ. Press, 1988).
38. Gouda, N., Sasaki, M. & Suto, Y. *Astrophys. J.* **341**, 557–574 (1989).
39. Sugiyama, N., Gouda, N. & Sasaki, M. *Astrophys. J.* **365**, 432–438 (1990).
40. Yokoyama, J. & Suto, Y. *Astrophys. J.* **379**, 427–439 (1991).
41. Anninos, P., Matzner, R. A., Tuluie, R. & Centrella, J. *Astrophys. J.* **382**, 71–78 (1991).
42. Efsthathiou, G., Sutherland, W. J. & Maddox, S. J. *Nature* **348**, 705–707 (1990).
43. Kaiser, N. & Stebbins, A. *Nature* **310**, 391–393 (1984).
44. Brandenberger, R. H. & Turok, N. *Phys. Rev. D* **338**, 2182–2187 (1986).
45. Stebbins, A. *Astrophys. J.* **327**, 584–614 (1988).
46. Bouchet, F. R., Bennett, D. P. & Stebbins, A. *Nature* **335**, 410–414 (1988).
47. Kashlinsky, A. & Jones, B. J. T. *Nature* **349**, 753–760 (1991).
48. Thompson, K. L. & Vishniac, E. A. *Astrophys. J.* **313**, 517–522 (1987).
49. Scaramella, R. *Astrophys. J.* **346**, 607–612 (1989).
50. Berstinger, E., Gorski, K. M. & Dekel, A. *Nature* **345**, 507–508 (1990).
51. Martinez-Gonzalez, E. & Sanz, J. L. *Mon. Not. R. astr. Soc.* **247**, 463–478 (1990).

LETTERS TO NATURE

Extreme gas pressures in the galactic bulge

David N. Spergel* & Leo Blitz†

* Princeton University Observatory, Princeton, New Jersey 08544, USA

† Astronomy Department, University of Maryland, College Park, Maryland 20742, USA

MOLECULAR gas in the innermost regions of the Galactic Centre must be in a very different state from the interstellar medium in the solar neighbourhood. Diffuse CS emission is seen only at the Galactic Centre^{1,2}, the emissivity of CO is sixty times higher there than near the Sun³, and linewidths of individual molecular features are typically five to ten times broader than elsewhere in the Galaxy. We propose that all these characteristics, for which there has so far been no coherent explanation, reflect the existence of high gas pressure—about $5 \times 10^6 \text{ K cm}^{-3}$, or two and a half orders of magnitude higher than that in the solar neighbourhood—in the inner 500 parsecs of the Galaxy. This high pressure can be inferred from the presence of hot X-ray-emitting gas, which is also known to be present in the bulges of many other galaxies^{4,5}. The molecular gas in these bulges should thus resemble that in the centre of our own Galaxy.

The molecular gas in any part of the Milky Way bulge should be in hydrostatic equilibrium with the galactic gravitational potential. Thus, the one-dimensional cloud–cloud velocity dispersion, σ , can be determined from the gas scale height, h_g , and the stellar density, ρ_*

$$\sigma = \alpha h_g \sqrt{2\pi G \rho_*} \quad (1)$$

since the gas scale height is much less than the stellar scale height. For a uniform disk, $\alpha = 1$, for a uniform sphere, $\alpha = \sqrt{2/3}$, and for a sphere with $\rho \propto r^{-2}$, $\alpha = \sqrt{2}$. K-band (2.2- μm) observa-

tions made with the infrared telescope flown aboard the space shuttle⁶ measure the stellar luminosity density. Assuming a mass-to-light ratio of 5 for the bulge⁷, these observations suggest $\rho_* \approx 25 M_\odot \text{ pc}^{-3}$, where $M_\odot$ is the solar mass, one degree from the centre. The molecular gas layer in the ¹³CO survey maps¹ has an apparent scale height of $\sim 0.2^\circ$, about 27 pc (throughout this paper, we assume that the distance to the Galactic Centre is 8 kpc). But this scale height could be smaller if the gas layer is tilted⁸. A very thin tilted gas disk projects into two discrete emission zones in the longitude–latitude plane at fixed radial velocity (see for example ref. 9). As these zones are not seen in the maps of ¹³CO in the Galactic Centre¹, the true scale height should be no less than a factor of 2 smaller than the apparent scale height. Using a stellar density of $25 M_\odot \text{ pc}^{-3}$ and a gas scale height of 20 pc in equation (1) yields $\sigma \approx 17 \text{ km s}^{-1}$, three times the local value^{10,11}.

The one-dimensional gas kinetic energy per unit volume

$$t = \Sigma_g \sigma^2 / h_g \quad (2)$$

is a kind of turbulent pressure represented by the cloud motions. Assuming hydrostatic equilibrium (equation (1))

$$t = 2\pi G \rho_* \Sigma_g h_g \quad (3)$$

If there is a diffuse gaseous component in the disk, this gas, too, must be in hydrostatic equilibrium, but in this case t represents the total gas pressure. Pressure equilibrium then requires that the total internal pressure of the various gas phases embedded in the diffuse gas be equal¹². Locally, this is indeed the case; the total gas pressure, $1.7 \times 10^4 \text{ K cm}^{-3}$, is close to the value expected from hydrostatic equilibrium¹³ and is comparable to the pressure in non-self-gravitating molecular clouds found at high galactic latitudes¹⁴.

If there is a diffuse gaseous component in the bulge, then t is expected to be much higher than in the disk. Observations of the 6.7-keV line of helium-like iron from the Ginga satellite¹⁵,

together with earlier observations of a diffuse X-ray background¹⁶⁻¹⁸, suggest that there is indeed a hot plenum of gas in the bulge extended over a substantial fraction of its volume. From the line strength and the continuum emission, Yamauchi *et al.*¹⁵ estimate that the gas temperature is ~ 10 keV and the gas density is $0.03\text{--}0.06\text{ cm}^{-3}$, implying a pressure of $\sim 3\text{--}6 \times 10^6\text{ K cm}^{-3}$. The r.m.s. velocity of this gas, $\sim 1,000\text{ km s}^{-1}$, implies that the hot gas is not in hydrostatic equilibrium but is escaping as a galactic wind. But if the unknown scale height of the gas is as large as ~ 1 kpc, the hot gas will be nearly hydrostatic (see below).

If the hot gas fills the inner few degrees of the bulge, as it seems to in the Ginga observations, then pressure equilibrium requires that the internal pressure within the gas clouds, the product of the density and the square of the turbulent velocity, be equal to the thermal pressure of the hot gas. For example, molecular gas at a density of 10^3 cm^{-3} , a typical density¹⁹ within clumps of local giant molecular clouds (GMCs), would have a one-dimensional turbulent velocity dispersion, v_{turb} , of $\sim 5\text{ km s}^{-1}$ (12 km s^{-1} full width at half maximum), several times higher than typical local values¹⁹⁻²². This is just what is observed in the maps of ^{13}CO maps in the bulge¹. Estimates of the linewidths of a number of discrete features in their maps give mean values very close to 12 km s^{-1} . If a local gas cloud were subjected to the higher pressures of the bulge, then depending on the value of γ , the ratio of specific heats, some of the pressure would be manifested as higher density and some as higher velocity dispersion. Relatively diffuse molecular gas could be found at very high densities, and typical densities of 10^4 cm^{-3} would be reasonable. Thus, the high pressure can account for both the large linewidths detected in the ^{13}CO in the region of the Galactic Centre and the presence of widespread diffuse CS in the inner few degrees of the Galaxy.

The turbulent pressure represented by the cloud motions need not be in pressure equilibrium with the hot gas any more than the turbulent pressure of the motions of stars need be coupled to that of the interstellar medium. Nevertheless, a near equilibrium is observed. The H_2 surface density within two degrees of the Galactic Centre is 60 times¹ the local value²³ of $\sim 1.3 M_{\odot}\text{ pc}^{-2}$ the local value of $1.3 M_{\odot}/\text{pc}^2$, thus, the turbulent pressure of the cloud motions is in near equilibrium with that of the hot gas. This is probably an indication that both components are close to hydrostatic equilibrium even though the hot gas must be escaping. The scale height of the hot gas is therefore expected to be ~ 1 kpc, a result which can be checked by modelling the X-ray emission.

One of the outstanding problems related to the molecular clouds in the bulge is that despite the very high surface density of molecules, the star formation rate does not seem to be significantly enhanced and may in fact be depressed (compare Figs 1 and 3 of ref. 24 with Fig. 5 of ref. 23; see also ref. 25). The large turbulent velocities provide a pressure that impedes collapse and may explain the low star-formation efficiency at the centre. The Jeans mass of clumps²⁶⁻²⁸

$$M_J = 10 \frac{v_{\text{turb}}^4}{(G^3 P_0)^{1/2}} = 10^6 \left(\frac{v_{\text{turb}}}{5\text{ km s}^{-1}} \right)^4 \left(\frac{P_0}{5 \times 10^6\text{ K cm}^{-3}} \right)^{-1/2} M_{\odot} \quad (4)$$

exceeds the cloud mass for clump densities of 10^3 cm^{-3} , so only a small fraction of the Galactic Centre clouds may be unstable to gravitational collapse. M_J will decrease as the mean clump density increases, but $M_J \propto n^{-1/2}$, a rather weak dependence. It would therefore not be surprising if star formation proceeded differently in the high-pressure environment of the bulge from in disk GMCs. Processes suggested for cooling flows²⁹ and for elliptical galaxies³⁰ may be relevant for the bulge of the Milky Way.

The physical conditions in the molecular clouds in the bulge

must differ considerably from the GMCs in the disk. The disk GMCs are self-gravitating³¹, and the quiescent clumps within local GMCs are in hydrostatic and pressure equilibrium with the self-gravity of the clouds¹⁹. The external pressure in the interstellar medium, $\sim 10^4\text{ K cm}^{-3}$, is negligible compared with the internal pressure within a GMC, $\sim 10^5\text{ K cm}^{-3}$. In the bulge, on the other hand, the external pressure will dominate, and only the most massive and densest molecular clouds are likely to be self-gravitating; the others instead will be pressure-bound, completely changing the character of the clouds. The use of CO as a mass tracer for H_2 depends on having similar molecular clouds throughout the Milky Way. As the clouds in the bulge are likely to be so different from those in the disk, the apparent CO/ H_2 ratio obtained from observations of the disk must be recalibrated in the bulge and may require significant revision (a similar problem for disk GMCs is discussed in ref. 32).

Unlike the gas layer in the disk, the molecular cloud layer seems to be stable against gravitational collapse. Rapid rotation stabilizes the bulge gas layer. The epicyclic frequency, κ , can be estimated from either the terminal velocity curve for the molecular gas or from the K-band photometry of the bulge. Either approach yields an epicyclic frequency of $\sim 2 \times 10^{-14}\text{ s}^{-1}$ at 300 pc from the Galactic Centre. Using our earlier estimates of the gas surface density and velocity dispersion, we can determine the Toomre parameter of the molecular layer

$$Q = \frac{\sigma \kappa}{3.36 G \Sigma} \approx 10 \quad (5)$$

Because $Q \gg 1$, the gas layer is gravitationally stable, so mechanisms that rely on gravitational instability to form GMCs would not operate in the bulge.

The H I in the galactic bulge cannot be in a diffuse state similar to that inferred from 21-cm observations of the disk. The overall profile shape of bulge H I can be reasonably explained by the large-scale kinematics of orbital motion about the centre, but the high-velocity edges of the profiles suggest that the velocity dispersion of the gas is of the order of 10 km s^{-1} (ref. 8). This dispersion suggests that the typical density of the H I gas is $\sim 10^2\text{--}10^3\text{ cm}^{-3}$ everywhere in the bulge. So as not to produce too much gas along the line of sight, the H I must have a relatively low filling fraction along the line of sight, because the surface density of the H I is somewhat less than the local value³³. This can be accomplished if the H I is confined to relatively thin sheets, most probably at the surfaces of the molecular clouds in the bulge.

Locally, the magnetic field pressure is in rough equipartition with the gas pressure^{34,35}. If this is also true in the bulge, then the r.m.s. magnetic field strength, B , in the Galaxy would be $\sim 130\text{ }\mu\text{G}$. A measure of the magnetic field strength in the Galactic Centre can be obtained from the synchrotron emissivity. The synchrotron flux from the inner two degrees is enhanced by a factor of three over the background value³⁶. Because the path length through this region is only 1/30 of the total path length, this enhancement suggests that synchrotron emissivity, which measures $n_e B^2$, is enhanced by two orders of magnitude over typical disk values. If the cosmic-ray electron abundance in the bulge, n_e , is comparable to the local value, the synchrotron observations suggest that the magnetic field pressure in the bulge is in equilibrium with the gas pressure. Future measurements of the magnetic field pressure by Zeeman splitting of H I or thermal OH can test whether these higher magnetic pressures exist within the central region of the Milky Way.

The cosmic-ray pressure, on the other hand, is likely to be much less in the bulge molecular gas than either the turbulent or magnetic field pressures. Observations of diffuse high-energy gamma-rays measure the product of the cosmic-ray density and the gas density. Gamma-ray data from the COS-B satellite of the Galactic Centre region show a gross deficiency of high-energy γ -rays, suggesting that the cosmic-ray density in the bulge is an order of magnitude less than the local value³⁷ or

that there is very much less H_2 in the bulge than is suggested by CO observations. The CS observations of Bally *et al.*¹ suggest that the former is the correct interpretation. Thus, in the bulge, the cosmic-ray nucleons do not seem to make a significant contribution to the gas pressure. This is consistent with our interpretation that the synchrotron observations imply a cosmic-ray electron density close to the local value.

The inference that the gas in the central regions of the Galaxy is under higher pressure leads to several observational consequences. Most of these are predictions, but some are suggested by existing data. (1) It will not be possible to find any lines of ^{13}CO unambiguously associated with the Galactic Centre with linewidths narrower than $\sim 10 \text{ km s}^{-1}$, regardless of the resolution used to observe them. (2) The volume density of the $H I$ in the galactic bulge is at least 10^2 – 10^3 everywhere in the bulge, and most probably exists only in thin sheets around the molecular clouds. (3) Measurements of Zeeman splitting in OH and $H I$ should show field strengths of $\sim 100 \mu G$ in gas unambiguously associated with the bulge. Such gas can be identified by its large radial velocities in absorption against the Sgr A and other continuum sources. (4) The scale height of the X-ray emitting gas is ~ 500 – $1,000 \text{ pc}$. (5) In the disk of the Milky Way, the ratio of high-density molecules to CO mapped within a constant CO surface brightness contour should be proportional to the stellar surface density of the disk and should decrease with galactocentric radius. The reason for this is that in the disk, the diffuse $H I$ is the hydrostatic component with the highest pressure because of its large scale height and high surface density at all distances outside the bulge. Because the stellar surface density falls off much more rapidly than the scale height increases³⁸, the total gas pressure must fall monotonically with galactic radius. (6) For a given ionizing flux, $H II$ regions should have higher emission measures (higher surface brightness) in the inner Galaxy than in the outer Galaxy; the emission measure scales as the square of the ambient gas density. In fact, the brightest $H II$ regions are seen in the molecular ring of the Galaxy where the gas pressures should have their highest values in the disk³⁹. (7) Supernova remnants in the galactic bulge should spend much less time in the 'snowplough' phase. Thus, there should be few large old remnants and there should be a much higher efficiency in transferring the kinetic energy of the supernova to the interstellar medium.

These arguments should be valid not only in the bulge of the Milky Way, but also in bulges of other galaxies and in elliptical galaxies. If there is an ambient hot, diffuse component filling these galactic bulges, then hydrostatic equilibrium and the deep potentials of these galaxies imply that the gas in the central regions of galaxies is under very high pressure⁴⁰. X-ray observations show that spirals and ellipticals frequently have such hot, high-pressure gas in their nuclei⁴¹. This leads to additional predictions. (1) The molecular gas in the bulge regions of galaxies should have a higher ratio of molecules such as CS, HCN and HCO^+ , relative to CO and ^{13}CO , than is found in their disks. In fact, high ratios have recently been found in the centres of many galaxies^{42,43}, suggesting that this picture is correct, but it remains to be shown that these ratios are significantly higher than is found in the disks of the galaxies, and that the observed ratios are correlated with central pressure. (2) The ratio of high-density molecules to CO should be a function of Hubble type and hot gas pressure. Dwarf irregular galaxies should have small abundances of high-density molecules relative to CO. In spite of the relatively weak lines of CO found in giant elliptical galaxies, they should have among the highest ratio of high-density molecules to CO if they have extended X-ray emission. If individual clouds mapped with interferometers can be observed, these clouds should have relatively broad lines commensurate with the large ambient pressures for the gas. (3) The inner regions of the molecular disk in the Cen A host galaxy NGC 5128 should have high internal velocity dispersions and high dense gas molecular abundances. $\square$

Received 17 January; accepted 11 May 1992.

- Bally, J., Stark, A. A., Wilson, R. W. & Henkel, C. *Astrophys. J. Suppl.* **65**, 13–82 (1987).
- Bally, J., Stark, A. A., Wilson, R. W. & Henkel, C. *Astrophys. J.* **324**, 223–247 (1988).
- Sanders, D. B., Solomon, P. M. & Scoville, N. Z. *Astrophys. J.* **276**, 182–203 (1984).
- Forman, W., Schwarz, J., Jones, C., Liller, W. & Fabian, A. C. *Astrophys. J.* **234**, L27–L31 (1979).
- Nansen, P. E. J., Stewart, G. C. & Fabian, A. C. *Mon. Not. R. Astr. Soc.* **208**, 185–195 (1984).
- Kent, S. M., Dame, T. M. & Fazio, G. *Astrophys. J.* **378**, 131–138 (1991).
- Faber, S. M. & Gallagher, J. S. A. *Rev. Astr. Astrophys.* **17**, 135–188 (1979).
- Liszt, H. S. & Burton, W. B. *Astrophys. J.* **236**, 779–797 (1980).
- Brinks, E. & Shane, W. W. *Astr. Astrophys. Suppl.* **55**, 179–257 (1984).
- Stark, A. A. *Astrophys. J.* **281**, 624–633 (1984).
- Magnani, L., Blitz, L. & Mundy, L. *Astrophys. J.* **295**, 402–421 (1985).
- Spitzer, L. *Astrophys. J.* **124**, 20–34 (1956).
- Bloemen, J. B. G. M. *Astrophys. J.* **322**, 694–705 (1987).
- Blitz, L. in *Molecular Clouds* (eds James, R. A. & Millar, T. J.) 49–68 (Cambridge Univ. Press, 1991).
- Yamauchi, S. *et al. Astrophys. J.* **365**, 532–538 (1991).
- Watson, M. G., Willingale, R., Grindlay, J. E. & Hertz, P. *Astrophys. J.* **250**, 142–154 (1981).
- Skinner, G. K. *et al. Nature* **330**, 544–547 (1987).
- Kawai, N. *et al. Astrophys. J.* **330**, 130–141 (1988).
- Blitz, L. in *The Physics of Star Formation and Early Stellar Evolution* (eds Lada, C. J. & Kylafis, N. D.) 3 (Kluwer, Dordrecht, 1991).
- Carr, J. S. *Astrophys. J.* **323**, 170–178 (1987).
- Loren, R. B. *Astrophys. J.* **338**, 902–924 (1989).
- Lada, E. A., Bally, J. & Stark, A. A. *Astrophys. J.* **368**, 432–444 (1991).
- Dame, T. M. *et al. Astrophys. J.* **322**, 706–720 (1987).
- Sodrowski, T. J., Dwek, E., Hauser, M. G. & Kerr, F. J. *Astrophys. J.* **322**, 101–122 (1987).
- Güsten, R. in *The Center of the Galaxy* (ed. Morris, M.) 89–106 (Kluwer, Dordrecht, 1989).
- Spitzer, L. *Diffuse Matter in Space* (Wiley, New York, 1968).
- Tomisaka, K., Ikeuchi, S. & Nakamura, T. *Astrophys. J.* **335**, 239–262 (1988).
- Bertoldi, F. & McKee, C. *Astrophys. J.* (in the press).
- White, R. E. & Sarazin, C. L. *Astrophys. J.* **318**, 612–620 (1987).
- Jura, M. *Astrophys. J.* **212**, 634–636 (1977).
- Solomon, P. M., Rivolo, A. R., Barrett, J. & Yahil, A. *Astrophys. J.* **319**, 730–741 (1987).
- Maloney, P. *Astrophys. J.* **348**, L9–L12 (1990).
- Burton, W. B. in *Galactic and Extragalactic Radio Astronomy* (eds Verschuur, G. L. & Kellerman, K. I.) 295–358 (Springer, Berlin, 1988).
- Rand, R. J. & Kulkarni, S. R. *Astrophys. J.* **343**, 760–772 (1989).
- Myers, P. C. in *Molecular Clouds* (eds James, R. A. & Millar, T. J.) 133–139 (Cambridge Univ. Press, 1991).
- Haslam, C. G. T., Salter, C. J., Stoffel, H. & Wilson, W. E. *Astr. Astrophys. Suppl.* **47**, 1–143 (1982).
- Blitz, L., Bloemen, J. B. G. M., Hermsen, W. & Bania, T. M. *Astr. Astrophys.* **143**, 267–273 (1985).
- Merrifield, M. *Astr. J.* (in the press).
- Altenhoff, W., Downes, D., Pauls, T. & Schraml, J. *Astr. Astrophys. Suppl.* **35**, 23–54 (1978).
- Mathews, W. G. & Baker, J. *Astrophys. J.* **170**, 241–259 (1971).
- Fabbiano, G., Kim, D.-W. & Trinchieri, G. *Astrophys. J. Suppl.* (in the press).
- Henkel, C., Baan, W. A. & Mauersberger, R. *Astr. Astrophys. Rev.* **3**, 47–90 (1990).
- Solomon, P. M., Downes, D. & Radford, S. J. E. *Astrophys. J.* **387**, L55–L60 (1972).

ACKNOWLEDGEMENTS. We thank E. Boldt, W. B. Burton, H. Liszt, J. Nousek, R. Ramaty, K. Weiler and C. Heiles for discussions. We also thank C. McKee for critically reading the manuscript. L.B. is supported by the NSF and NASA. D.N.S. is supported by a Sloan Fellowship and the NSF.

^{14}C activity of dissolved organic carbon fractions in the north-central Pacific and Sargasso Sea

James E. Bauer*, Peter M. Williams†
& Ellen R. M. Druffel‡

* Department of Oceanography, Florida State University, Tallahassee, Florida 32306, USA

† Marine Research Division, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA

‡ Department of Chemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

RADIOCARBON measurements of dissolved organic carbon (DOC) oxidizable by ultraviolet irradiation (DOC_{ox}) yielded apparent ages of $\sim 6,000 \text{ yr}$ in the deep waters of the oligotrophic north-central Pacific gyre¹. Recent reports of a potentially larger pool of DOC as measured by high-temperature catalytic combustion (DOC_{htc}) using discrete injections of sea water^{2,3} have led to speculation that 'younger', more recently produced DOC could contribute significantly to overall oceanic organic carbon fluxes, owing to its suspected greater biological lability^{4–6}. Here we present a comparison of $\Delta^{14}C$ (the deviation in parts per thousand from the ^{14}C activity of nineteenth-century wood)⁷ of the DOC_{htc} , DOC_{ox} , and humic substances in profiles from the oligotrophic north-central Pacific and Sargasso Sea. For each ocean, the $\Delta^{14}C$ values of all three fractions are remarkably similar, yielding no

evidence for a component of DOC that is cycled through the system on timescales shorter than several thousands of years. We observe an age difference between the two oceans of $\sim 2,000$ yr for the deepest DOC, which can largely be accounted for by differences in the $\Delta^{14}\text{C}$ of the DOC sources to the deep basins, and by the different deep-water circulation patterns and transit times in the two oceans.

Here we use the term DOC to mean that fraction of the organic carbon which passes a Whatman GF/C glass-fibre filter (nominal pore size $1.0\ \mu\text{m}$), thus including colloidal organic carbon and traces of carbon associated with small bacteria, viruses and organic particles. A total of 27 analyses of sea water from depths of 3 to 5,700 m in the north-central Pacific (Eve-1 cruise; $31^\circ 00' \text{N}$, $159^\circ 00' \text{W}$; June 1987) and 34 analyses from 3 to 4,450 m in the Sargasso Sea (Hydros-6 cruise; $31^\circ 50' \text{N}$, $63^\circ 30' \text{W}$; May 1989) were made on samples collected with clean, stainless steel Gerard barrels and filtered directly into clean 4-l bottles which were frozen at -20°C . Before analysis, samples were thawed, acidified to pH 2.5 with H_3PO_4 and

sparged with carbon-free N_2 or O_2 to remove inorganic carbon. The DOC in the sea water was then oxidized by either standard ultraviolet irradiation ($1,200\ \text{W}$ for $6\ \text{h}$)¹, high-energy ultraviolet irradiation ($2,400$ – $3,600\ \text{W}$ for 2 – $3\ \text{h}$), or continuous-injection HTC oxidation using Co/CoO-coated alumina or 2% (by weight) Pt-coated zeolite catalysts⁸. Separate subsamples were also analysed by HTC using discrete, $100\text{-}\mu\text{l}$ injection volumes. Humic substances (humic plus fulvic acids) were obtained by sequentially extracting sea water acidified with HCl to pH 2 on XAD-2 and XAD-4 macroreticular resins and collecting the base-soluble ($0.1\ \text{M NaOH}$) fraction of the eluate^{9,10}. The recovered, desalted material was then oxidized to CO_2 by sealed-tube combustion¹¹. All samples were analysed for $\Delta^{14}\text{C}$ by accelerator mass spectrometry (AMS).

The efficacy of continuous-injection HTC combustion for quantitatively oxidizing DOC and allowing accurate determination of its ^{14}C content was demonstrated by redissolving humic substances, isolated from 50 and 3,200 m at the Sargasso Sea site, in ultra-high-purity artificial seawater at concentrations

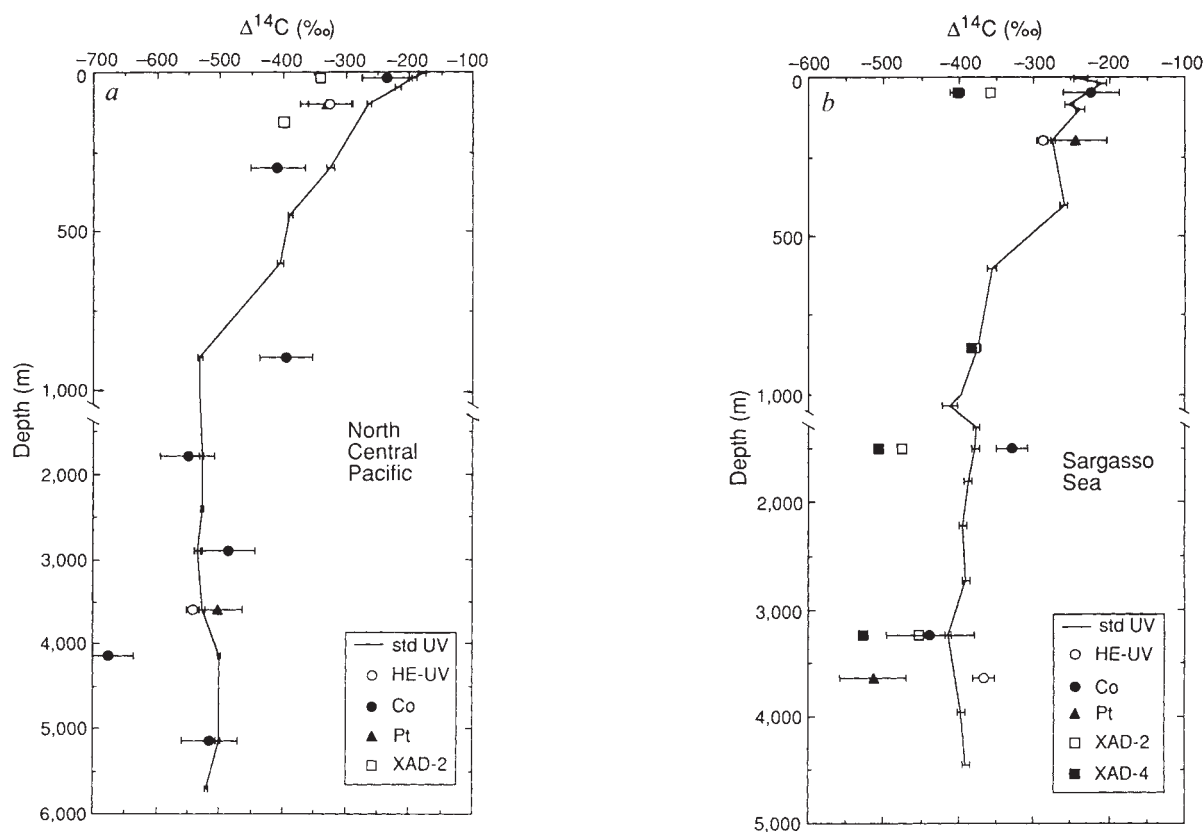


FIG. 1 Profiles of $\Delta^{14}\text{C}$ values of DOC measured for sea water from (a) the north-central Pacific ($31^\circ 00' \text{N}$, $159^\circ 00' \text{W}$; June 1987) and (b) the Sargasso Sea ($31^\circ 50' \text{N}$, $63^\circ 30' \text{W}$; May 1989). Values of $\Delta^{14}\text{C}$ for DOC_{UV} from the Pacific were previously reported²⁴. All $\Delta^{14}\text{C}$ values are corrected to a $\delta^{13}\text{C}$ value of -25‰ (ref. 7). The $\delta^{13}\text{C}$ values of DOC_{UV} were determined directly on sample CO_2 , whereas that of DOC_{HTC} was assumed to be -21‰ . Horizontal bars represent total cumulative error, including that for AMS counting statistics, blank correction and replicate analysis. Errors for $\Delta^{14}\text{C}$ values of DOC_{HTC} are large because of subtraction of experimental blanks and their associated $\Delta^{14}\text{C}$. Standard ultraviolet (std UV) oxidations followed previous methodology¹. High-energy ultraviolet (HE-UV) oxidations used similar techniques but with smaller sample volumes ($650\ \text{ml}$) irradiated for shorter periods of time ($2\ \text{h}$) at higher energies ($2,400$ – $3,600\ \text{W}$). The system blank for all ultraviolet oxidations was $<2\%$ of the sample size. The HTC oxidations consisted of continuous injection of seawater samples onto either Co/CoO-on-alumina or 2% Pt-on-zeolite catalysts at 900°C , using ultra-high-purity (UHP) oxygen as the carrier gas²⁹. Each HTC oxidation of $\sim 100\ \text{ml}$ of sea water required ~ 3 – $3.5\ \text{h}$. The blank due to carrier-gas flow (no water injection) averaged $22 \pm 8\%$ of total carbon collected during a sample run.

The average blank attributable to injection of water during HTC oxidations was $1.4 \pm 2.5\ \mu\text{g C h}^{-1}$ and was attained by continuously injecting UHP distilled water through the system for 12 – $18\ \text{h}$ (total volume ~ 400 – $500\ \text{ml}$) before sample analysis. The mean $\Delta^{14}\text{C}$ value of blanks for carrier gas only over all runs was $-441 \pm 167\text{‰}$ compared with a mean $\Delta^{14}\text{C}$ value of $-430 \pm 200\text{‰}$ for continuous injection of UHP distilled water. The CO_2 evolved from each oxidation was collected either cryogenically for ultraviolet oxidations or with a synthetic zeolite molecular sieve for HTC oxidations²⁵, purified and quantified on a vacuum extraction line and flame-sealed in a 6-mm Pyrex tube. The $\Delta^{14}\text{C}$ measurements were made at Lawrence Livermore National Laboratory or University of Arizona AMS facilities following the conversion of sample CO_2 to graphite targets^{26,27}. The $\Delta^{14}\text{C}$ values of humic substances isolated from the north-central Pacific were determined previously¹² (K. Meyers-Schulte and J. Hedges, personal communication) and are presented for comparison. In the Sargasso Sea, the XAD-2 isolates of humic substances are defined as having higher average molecular weight and hydrophobicity than the XAD-4 isolates. Note difference in depth scales between sites and change in scale at $1,000\ \text{m}$.

TABLE 1 Recoveries of carbon from north-central Pacific and Sargasso Sea water

Depth (m)*	HTC		Ultraviolet		HTC Discrete injection§	
	DOC _{htc} (μM)†	Δ ¹⁴ C (‰)‡	DOC _{uv} (μM)†	Δ ¹⁴ C (‰)‡	Depth (m)	DOC (μM)
North-central Pacific						
20	82	-236 ± 39	65	-195 ± 6	20	218 ± 6
100	71	-331 ± 41	53	-264 ± 4	100	206 ± 4
300	85 ± 16	-409 ± 43	53	-325 ± 5	300	151 ± 3
896	83	-395 ± 41	42	-532 ± 3	900	90 ± 3
1,788	57 ± 2	-551 ± 44	35	-529 ± 3	1,769	n.d.
2,891	76	-486 ± 41	35	-535 ± 4	3,174	n.d.
3,586	63	-503 ± 40	34	-528 ± 4	3,631	110 ± 3
4,146	67	-675 ± 38	36	-501 ± 3	4,227	112 ± 3
5,150	85	-516 ± 45	36	-502 ± 3	5,227	112 ± 3
Sargasso Sea						
50	87 ± 11	-224 ± 36	62	lost	50	130 ± 12
200	101	-245 ± 41	54	-276 ± 5	200	114 ± 3
1,510	61	-370 ± 20	43	-378 ± 5	1,510	98 ± 6
3,237	82 ± 6	-438 ± 59	43	-414 ± 5	3,237	103 ± 2
3,638	66	-514 ± 44	44	lost	3,638	102 ± 2

* Depths listed are those at which both continuous-injection HTC and ultraviolet oxidations were done.

† Concentrations determined by collection, vacuum-line purification and quantification of CO₂ derived from either continuous-injection HTC or standard ultraviolet oxidations. For samples oxidized by continuous-injection HTC oxidation, we used a Co/CoO-on-alumina catalyst, except for those designated by (||), for which we used 2% Pt-on-zeolite catalyst. Measurement error (±1σ) for concentration determinations was ±1 μM. Errors associated with DOC concentrations are ±1σ of the mean of duplicate oxidations.

‡ Δ¹⁴C errors are ±1σ resulting from counting statistics, sample δ¹³C and blank Δ¹⁴C corrections; || indicates average ±1σ of replicate oxidations. For samples oxidized by continuous-injection HTC we used a Co/CoO-on-alumina catalyst, except for those designated by (||), for which we used 2% Pt-on-zeolite catalyst.

§ Discrete-injection DOC measurements made using either 3% Pt-on-alumina (north-central Pacific samples)¹² or Co/CoO-on-alumina (Sargasso Sea samples) catalysts²³. Values for the north-central Pacific were previously reported¹² for the Alcione-5 cruise to the same site, resulting in slight depth differences. Values for the Sargasso Sea were calculated from standard CO₂ and glucose calibration curves which were corrected by appropriate standard water (ultra-high-purity double-distilled) blank values. Errors associated with DOC concentration values are ±1σ of the mean of 3–5 replicate injections. n.d., not determined.

representative of the 'total' DOC and treating the solutions to the same procedure as natural seawater samples. Recoveries of the humic substances were all 100 ± 2% of the known amounts added, and Δ¹⁴C values were identical to those determined by sealed-tube combustion of the solid humics (-374 ± 21% by HTC and -387 ± 26% by STC for 50 m; -509 ± 30% by HTC and -495 ± 30% by STC for 3,200 m).

We investigated the possibility that the Δ¹⁴C of DOC_{uv} is biased because of selective oxidation of organic matter. First, serial ultraviolet oxidations of samples from 20 and 1,800 m in the north-central Pacific showed no consistent differences in Δ¹⁴C values of the material that was oxidized over each interval. Secondly, high-energy ultraviolet irradiation, which oxidized 4–18% more DOC than the standard ultraviolet irradiation, gave Δ¹⁴C values similar to those obtained by the standard ultraviolet irradiation (Fig. 1a and b). Both results indicate that little fractionation of ¹⁴C of DOC_{uv} occurs as a function of DOC_{uv} recovery.

The Δ¹⁴C of DOC_{htc} strongly resembled the more extensive profiles for DOC_{uv} in both the north-central Pacific and Sargasso Sea (Fig. 1a and b). The Δ¹⁴C of DOC_{uv} from the north-central Pacific (Fig. 1a) ranged from -179‰ at 3 m to -535‰ at 2,900 m, similar to values obtained for sea water collected at this site in

1985 (ref. 1). The range of Δ¹⁴C values for DOC_{htc} (excluding the value at 4,146 m) was -236‰ at 20 m to -551‰ at 1,800 m. Nearly all of the change in Δ¹⁴C of the DOC fractions occurred in the upper 1,000 m. The range in Δ¹⁴C values from 3 m to 900 m was 353‰; from 900 m to 5,700 m it was only 34‰. The significantly greater Δ¹⁴C value of the sample from 900 m (O₂ minimum zone) measured by HTC oxidation may be indicative of a biologically refractory fraction having a younger apparent age, whereas the low Δ¹⁴C value at 4,146 m seems to be anomalous. For all samples shallower than 300 m, the DOC_{htc} had slightly lower Δ¹⁴C values (>1σ but <2σ) than the DOC_{uv} (Fig. 1a).

The Δ¹⁴C values of DOC_{htc} for Sargasso Sea water were also, with one exception (3,640 m, Pt-zeolite catalyst), similar to the DOC_{uv} Δ¹⁴C values (Fig. 1b). The Δ¹⁴C profiles of DOC_{uv} from the Sargasso Sea were notably different from those for the north-central Pacific. At 20 m, the Δ¹⁴C value was 15‰ lower in the Sargasso Sea because of smaller amounts of recently fixed DOC in the surface waters. At all depths greater than 20 m, however, Δ¹⁴C values of DOC_{uv} were significantly greater in the Sargasso Sea. The magnitude of the offset between Atlantic and Pacific Ocean waters averaged -41‰ from 50 m to 600 m and -139‰ from 900 m to ~3,400 m. As in the north-central Pacific, nearly the entire gradient in DOC Δ¹⁴C values occurred in the top 1,000 m and can be attributed to the presence of recently fixed carbon ultimately derived from planktonic production in surface waters.

The recoveries of DOC using continuous-injection HTC oxidation were consistently greater than, and varied relative to, the standard ultraviolet oxidations (Table 1). For both sites, continuous-injection HTC oxidation recovered 25–40% more DOC from shallow waters (20–100 m) and 83 ± 28% (N = 11) more from deep waters (>100 m) than did ultraviolet oxidation. The recoveries by continuous-injection HTC relative to discrete-injection HTC could be due to several factors including incomplete oxidation, loss of DOC during freezing and storage, or adsorption of hydrophobic components of the DOC in near-surface samples. The similarity of Δ¹⁴C values for DOC_{htc} and DOC_{uv} spanning this range of recoveries suggests that these fractions are characteristic of the mean Δ¹⁴C values for the average bulk DOC pool, especially in the deep sea.

The δ¹³C values of representative DOC_{htc} samples from 200 and 3,600 m in the Sargasso Sea were -22.5 ± 0.4‰ (N = 3) and -21.5 ± 0.4‰ (N = 3), respectively, indicating that DOC_{htc} is marine in nature and that no significant isotope fractionation occurred during HTC oxidation. For Sargasso Sea samples, the mean δ¹³C value of DOC_{uv} was -20.9 ± 0.3‰ (N = 18) and the average value over all depths for the north-central Pacific was -21.2 ± 0.5‰ (N = 17), identical to previously reported values for this location¹. This indicates that the DOC_{uv} fraction in both oceans is the same with respect to δ¹³C and is also of marine origin.

The Δ¹⁴C values of humic material from north-central Pacific sea water (Fig. 1a) were previously found to be -342‰ (20 m) and -400‰ (180 m), respectively¹² (K. Meyers-Schulte and J. Hedges, personal communication). In the Sargasso Sea, the Δ¹⁴C values of humic isolates were significantly lower than DOC_{uv} (Fig. 1b) except at 850 m (O₂ minimum), where Δ¹⁴C values of the XAD-2 and XAD-4 isolates (XAD-4 is a lower-molecular-weight, more-hydrophilic fraction than that isolated on XAD-2) and DOC_{uv} were identical, indicating that these fractions of the DOC all cycle on similar timescales at this depth.

In both sites, total concentrations of DOC measured by continuous-injection HTC and ultraviolet oxidations, both of which use direct collection and quantification of CO₂, were less than those measured by discrete-injection HTC using infrared analysis of CO₂ (Table 1). Assuming that a portion (~30%) of the nonhumic DOC pool was not combusted and was unfractionated with respect to ¹⁴C by continuous-injection HTC oxidation (a contention supported by the data), we may calculate the Δ¹⁴C

values of a hypothetical 'total' DOC pool, assuming that the discrete-injection DOC_{htc} values are accurate. (Currently this accuracy is an open question²⁸.) We further assume that this additional fraction is younger and has a post-bomb $\Delta^{14}\text{C}$ value of +120‰, similar to that of the dissolved inorganic carbon (DIC) in the mixed layer¹³. The calculated $\Delta^{14}\text{C}$ values of this 'total' DOC pool for the shallowest and deepest samples measured using Co/CoO HTC are -129‰ (20 m) and -325‰ (5,150 m) in the north-central Pacific and -121‰ (50 m) and -271‰ (3240 m) in the Sargasso Sea. If recoveries of DOC in the shallow samples were only 50% of the DOC measured by discrete injection (Table 1), then the $\Delta^{14}\text{C}$ value of the total shallow-water DOC pool would increase to -60‰ (apparent age 500 yr BP) in the north-central Pacific and -52‰ (apparent age 430 yr BP) in the Sargasso Sea. However, this poses two paradoxes: first, that the hypothetical extra, 'modern' DOC is impervious to either ultraviolet or continuous-injection HTC oxidation, and second, that ~30% of the 'total' DOC in the deep ocean is uniformly modern.

As the DOC ages are mean values, it can be calculated from a two-age component model¹⁴ that 40% of the DOC in the deep Pacific and 55% in the deep Atlantic could have a post-bomb radiocarbon signature ($\Delta^{14}\text{C} = +120\text{‰}$), provided that the remainder of the DOC is 'dead' ($\Delta^{14}\text{C} = -1,000\text{‰}$). Discussions have thus far implied that $\Delta^{14}\text{C}$ values associated with operationally defined fractions represent discrete groups or types of DOC. Although this possibility cannot be unequivocally ruled out, DOC in ocean waters may be more realistically described as a continuum of material of all apparent ages ranging from modern to fossil. Therefore, we cannot preclude the input to the deep ocean of a significant proportion of DOC derived from incorporation of pre-bomb ($\Delta^{14}\text{C} = -50\text{‰}$) to post-bomb ($\Delta^{14}\text{C} = +120\text{‰}$) DIC into the DOC pool via carbon fixation in surface waters. In fact, in both sites the presence of a DOC component of higher $\Delta^{14}\text{C}$ than average bulk DOC oxidized by ultraviolet or HTC can be deduced by simple mass balance of this bulk DOC and the DOC represented by the humic isolates, which are completely oxidized by both methods (Fig. 1a and b). A better understanding of the chemical make-up of deep-ocean DOC, its heterotrophic utilization, and the $\Delta^{14}\text{C}$ distributions of refractory and biologically usable DOC, as well as further direct collections of CO_2 from DOC oxidation, will help resolve these apparent discrepancies.

The combined DOC_{htc} and DOC_{uv} $\Delta^{14}\text{C}$ data sets indicate average minimum apparent ages for DOC in north-central Pacific and Sargasso Sea near-surface sea water (roughly 0–900 m) of at least 3,100 and 2,500 yr BP, respectively. If we account for the certain presence of post-bomb carbon in the DOC, these ages are greater still. The average apparent ages of DOC in the deep ocean below 900 m (5,900 and 4,100 yr BP in the north-central Pacific and Sargasso Sea, respectively) are also minimum estimates given the likely presence of $\leq 10\%$ post-bomb DOC ($\Delta^{14}\text{C} \approx +120\text{‰}$) in the deep sea. This old DOC fraction must therefore represent biologically and chemically resistant background material in oligotrophic regions and may account for previous observations of extremely low rates of heterotrophic respiration and DOC utilization in the deep ocean^{15–17}. The fact that the apparent ^{14}C age of DOC in the deep Sargasso Sea is ~2,000 yr younger than that in the north-central Pacific suggests that a high proportion of the total DOC is transported along with deep waters in the world ocean, which have transit times of ~1,400 yr from the North Atlantic to the North Pacific^{18,19,29}.

The ultimate sinks of organic carbon produced in the euphotic zone, including both dissolved and particulate forms, are removal to the sediments^{20,21} and heterotrophic respiration²². Processes that bypass these sinks will result in a net recycling of organic carbon and an increase in its residence time. Both the extent of recycling and the residence time of DOC will be directly proportional to its efficiency of conversion to biomass

and inversely proportional to its mineralization rate, neither of which is well understood in the deep ocean. Even at relatively high conversion efficiencies (say, ~70%), a diminishing proportion of organic carbon will be recycled, especially at rapid rates of utilization (years to decades). With respect to the ~30% of the 'total' DOC from deep ocean waters that might not be oxidized by continuous-injection HTC oxidation, there is no evidence that this fraction is modern and heterotrophically labile. The great apparent ages of the DOC fractions seen here confirm that most of the DOC in the deep ocean is remineralized extremely slowly. □

Received 27 November 1991; accepted 30 April 1992.

- Williams, P. & Druffel, E. *Nature* **330**, 246–248 (1987).
- Sugimura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
- Martin, J. & Fitzwater, S. *Nature* **356**, 699–700 (1992).
- Toggweiler, J. in *Productivity of the Ocean: Past and Present* (eds Berger, W., Smetacek, V. & Wefer, G.) 65–83 (Wiley, New York, 1989).
- Williams, P. & Bruffel, E. *Oceanogr. Mag.* **1**, 14–17 (1988).
- Kirchman, D., Suzuki, Y., Garside, C. & Ducklow, H. *Nature* **352**, 612–614 (1991).
- Stuiver, M. & Polach, H. *Radiocarbon* **19**, 355–363 (1977).
- Bauer, J., Ocelli, M., Williams, P. & McCaslin, P. *Mar. Chem.* (in the press).
- Thurman, E. & R. Malcolm, *Envir. Sci. Technol.* **15**, 463–466 (1981).
- Aiken, G. in *Humic Substances in Soil, Sediment and Water* (eds Aiken, G., McKnight, D., Wershaw, R. & MacCarthy, P.) 363–386 (Wiley, New York, 1985).
- Sofer, Z. *Analyt. Chem.* **52**, 1389–1391 (1980).
- Druffel, E., Williams, P. & Suzuki, Y. *Geophys. Res. Lett.* **16**, 991–994 (1989).
- Druffel, E. *et al.* *Radiocarbon* **31**, 523–532 (1989).
- Mantoura, R. & Woodward, E. *Geochim. cosmochim. Acta* **47**, 1293–1309 (1983).
- Keys, A., Christensen, E. & Krogh, A. *J. mar. Biol. Ass. UK* **29**, 181–196 (1935).
- Rakestraw, N. *J. mar. Res.* **3**, 259–263 (1947).
- Barber, R. *Nature* **220**, 274–275 (1968).
- Broecker, W. *Chemical Oceanography* (Harcourt Brace, Jovanovich, New York, 1974).
- Stuiver, M., Quay, P. & Ostlund, H. *Science* **219**, 849–854 (1983).
- Emerson, S. & Hedges, J. *Paleoceanography* **3**, 621–634 (1988).
- Hedges, J. *Mar. Chem.* (in the press).
- Smith, S. & Mackenzie, F. *Global biogeochem. Cycles* **1**, 187–198 (1987).
- Bauer, J., Williams, P., Druffel, E. & Suzuki, Y. *Eos* **71**, 154 (1990).
- Druffel, E. & Williams, P. *Nature* **347**, 172–174 (1990).
- Bauer, J., Williams, P. & Druffel, E. *Analyt. Chem.* **64**, 824 (1992).
- Vogel, J., Nelson, D. & Southon, J. *Radiocarbon* **29**, 323–333 (1987).
- Jull, A., Donohue, D., Hathaway, A., Linick, T. & Toolin, L. *Radiocarbon* **28**, 191–196 (1986).
- Williams, P. M. *US JGOFS News* **3**, 1 (1991).
- Druffel, E., Williams, P., Bauer, J. & Ertel, J. *J. geophys. Res.* (in the press).

ACKNOWLEDGEMENTS. We thank J. Hedges for comments on the manuscript; J. Vogel, J. Davis, S. Trumbore, D. Lloyd, J. Southon, and A. Jull, D. Donahue and L. Toolin for $\Delta^{14}\text{C}$ analyses; K. Robertson, J. Paduan, J. Ertel, S. Griffin and A. Witter for help with sample collection and technical assistance; M. Ocelli for the Pt catalyst; and the crew of the RV *Melville* for help with sample collection and ship logistics. Financial support from the NSF Chemical Oceanography Program.

Rapid coupling of sinking particle fluxes between surface and deep ocean waters

V. L. Asper*, W. G. Deuser†, G. A. Knauer* & S. E. Lohrenz*

* University of Southern Mississippi, Center for Marine Science, John C. Stennis Space Center, SSC, Mississippi 39529, USA

† Woods Hole Oceanographic Institution, Department of Chemistry, Woods Hole, Massachusetts 02543, USA

SETTLING particles are thought to be responsible for much of the transport of mass and energy from the upper ocean to the sea floor. Photosynthetic production by phytoplankton is a major source of these particles, either as phytoplankton biomass sinks directly¹ or as it is transformed into rapidly sinking forms such as aggregates^{2,3} and zooplankton faeces⁴. Because a variety of processes may act on sinking matter, however, it is not known to what extent fluxes of organic matter to the deep sea are coupled to processes at the ocean surface. Some studies have provided evidence for direct coupling^{2,5–7}, but transformation processes and advection exist which have the potential to modify the transmission of surface signals to the deep sea^{8–11}. If these mechanisms overwhelm surface production signals, seasonal and annual variations

in deep-sea geochemistry and biology would be controlled largely by lateral processes associated with ocean circulation rather than by surface processes. Here we report direct measurements of seasonal variations in upper-ocean primary production concurrent with particle fluxes measured at several depths ranging from the upper to the deep ocean in the Atlantic. We find that the productivity signal can be transferred rapidly to the deep sea by settling particles, yielding close temporal coupling between the surface and deep oceans.

Monthly measurements of primary production and near-surface particle flux were obtained between December 1988 and December 1990 at the oligotrophic Oceanic Flux Program/Bermuda Atlantic Time-series Study (OFP/BATS) site (31° 50' N, 64° 10' W), 80 km southeast of Bermuda. Primary production was determined by *in situ* ^{14}C incubations during each sampling cruise using trace-metal-free techniques^{12,13}. Upper ocean mass flux was measured monthly at depths of 150, 200, 300 and 400 m using multiple cylindrical sediment traps suspended on a floating array generally for a 72 h period¹⁴. Deep flux (500 m, 1,500 m and 3,200 m) was measured continuously, initially over 2-month and later over 2-week-intervals using single-cone traps^{15,16}.

Comparison of the data sets reveals a clear correspondence of the major peaks of primary production (Fig. 1a), surface particle flux (Fig. 1b–e), and deep ocean particle flux (Fig. 1f–h). Given the temporal resolution of our monthly sampling, it seems that particles associated with the pulses of primary production in March 1989, July 1989 and March/April 1990 settled at sufficiently high rates to be collected in sediment traps

throughout the water column at roughly the same time. Assuming a maximum delay of 30 days (that is, the interval of our discrete surface sampling), the average sinking speed of this material must exceed 100 m per day for it to arrive at all depths simultaneously (within the limits of our resolution). Seasonal variability in flux has been previously reported for deep moored traps (3,200 m) in the OFP/BATS area^{5,6,17–19}, in the Panama Basin²⁰ and in the Black Sea²¹. But most previous studies were only able to infer temporal correspondence of near-surface primary production and particle flux as complementary time series information on upper ocean biogeochemical processes was not available. An exception was the recently described relationship between surface-ocean pigment concentrations measured by satellite and deep-ocean carbon flux measured by sediment trap²². Our concurrent *in situ* measurements of production by autotrophs, export from the mixed layer and the arrival of particles in the deep sea show empirically the potential for direct and rapid coupling between these processes throughout the water column.

In addition to the differences in deployment configuration (floating array versus bottom-moored), the sampling regimen for the surface traps differed substantially from that of the deep traps, resulting in unequal temporal coverage. The floating traps sampled only 3 days per month (10% coverage) and primary production was measured one day per month (3% coverage). In contrast, the deep trap deployments were conducted consecutively to yield nearly continuous sampling over the entire comparison period. These differences in sampling intervals limit the certainty that can be associated with comparisons of surface and deep samples and could result in apparent disparities between the three data sets. Possible examples of these disparities are seen in cases where fluxes reaching the deep trap seem to lead surface production slightly in April 1990, although this may be a result of aliasing due to the discrete sampling applied in surface waters as opposed to the continuous sampling applied in the deep traps.

Short duration production events⁷ might also have resulted in discrepancies between the surface and deep signals. These events could easily be missed by the intermittent sampling conducted at the surface, but would be recorded by the deep traps due to the integrative nature of their sampling regime. During periods of nutrient limitation, storm events could have produced mixing which supplied nutrients to the surface water and resulted in enhanced productivity^{7,27}. In such cases, the surface samples would not necessarily be expected to correspond well with the continuously collected deep samples.

Close examination of these data reveals that although evidence for vertical particle settling is clear, a one-dimensional view of particle flux in the ocean may not be complete. The correspondence between the three data sets is driven by the seasonal cycle of phytoplankton growth and decline. During bloom periods (solid vertical lines in Fig. 1), particle fluxes decrease with depth as would be expected if particles are remineralized or disaggregated during settling. During non-bloom periods, the flux of particles shows an apparent mid-depth minimum with maximal values near the surface and at great depths (compare Fig. 1a–d with e–g). This profile, with a mid-water minimum in particle flux, is not unusual and has been observed in many locations, including the Black Sea²¹, Gulf of Mexico²³, Panama Basin²⁴ and the North Pacific^{12,25}, where additional mid-depth increases were seen in addition to surface and deep flux maxima^{2,12}. The origins of these mid-depth minima remain elusive, but a reasonable explanation invokes removal of material by decomposition, consumption and remineralization above the minimum and the incorporation of dissolved and/or suspended material to the sinking particle pool at depths below the minimum.

Another possible explanation for the mid-depth particle flux minimum relates to the 'statistical funnel' through which settling particles arrive in a sediment trap^{22,26,27}. Particles settling into

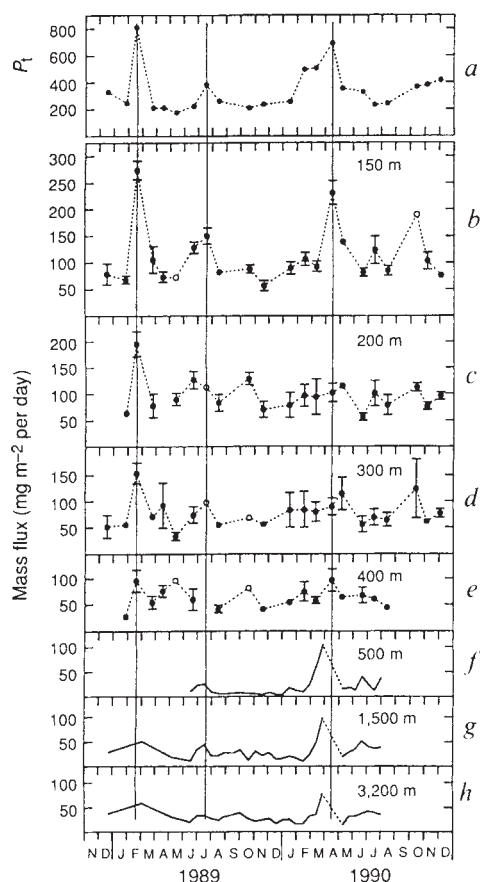


FIG. 1 Data from the OFP site near Bermuda comparing primary production (P_t in units of mg C m^{-2} per day) (a) with floating (b–e) and bottom-tethered sediment trap results (f–h). Error bars represent one standard error of the mean; open circles represent points where a single sample was available; dotted lines indicate sampling hiatuses. The seasonal signal of variations in production is evident in all samples at essentially the same time.

any trap fall along trajectories determined by their fall velocity and the ambient flow field. Deeper traps, stronger flows and smaller slow sinking particles result in larger 'funnel' openings at the sea surface. For our 1,500 m and 3,200 m traps, the funnel should be large enough to encompass regions to the north of the sampling site known to have higher productivity and particle export²². Particles exported from these regions may reach the deeper traps and enhance the magnitude of the signal. Further enhancement may also occur when aggregated matter sinking from directly above the deep traps interacts with suspended particles from distant sources, scavenges them from the water column and transmits them to the sediment²⁸.

The mid-depth flux minimum could also be due to the removal of material from the traps at the depth of the minimum due to hydrodynamic effects of water flowing around and over the trap openings. Over the last decade, a number of studies have reported variations in the efficiency of sediment traps as a result of differing current regimes in the trapping environment^{29–32}. However, if hydrodynamic conditions were so extreme as to remove such a major portion of the sample from the mid-depth traps, then it is likely that the temporal flux signal would have been masked by this process. If so, then trap collections (particle flux) would be correlated with changes in the flow of water past the trap openings, and would not correspond to the seasonal signal in primary productivity. The close correspondence between the trap measurements and independently determined primary productivity strongly suggests that this is not the case. Any hydrodynamic effects acting on the traps thus seem insufficient in this case to overwhelm the transmission of the productivity signal through the water column.

The overall correspondence between the primary production data and both the surface and the deep traps, supports the premise that the annual signal of surface productivity is transferred rapidly through the water column through sinking particles. This rapid link between the surface and deep ocean can result in significant seasonal and annual variations in the availability of energy to the benthic ecosystem. Thus, despite the absence of light or other seasonal physical changes, deep ocean biota experience seasonal environmental patterns which are closely coupled to those occurring in the surface ocean. The rapid transport of particles through the water column also produces a close spatial coupling between the surface and the sea floor, as rapidly sinking particles are deposited relatively near their production sites. Finally, the close correspondence between the three data sets indicates that the signals were recorded with small, or at least consistent, biases by our techniques. The net effect of environmental factors acting on our traps, such as hydrodynamics and swimmers³³, was insufficient to obscure the qualitative nature of this signal. □

26. Deuser, W. G., Muller-Karger, F. E. & Hemleben, C. *J. geophys. Res.* **93**, 6857–6862 (1988).
27. Siegel, D. A., Granata, T. C., Michaels, A. F. & Dickey, T. D. *J. geophys. Res.* **95**, 5305–5311 (1990).
28. Asper, V. L., Thesis, Woods Hole Oceanographic Institution, Massachusetts Institute of Technology (1986).
29. Brewer, P. G., Nozaki, Y., Spencer, D. W. & Fleer, A. P. *J. Mar. Res.* **38**, 703–738 (1980).
30. Butman, C. A., *J. Mar. Res.* **44**, 645–693 (1986).
31. Butman, C. A., Grant, W. D. & Stolzenbach, K. D. *J. Mar. Res.* **44**, 601–644 (1986).
32. Gardner, W. D. *J. Mar. Res.* **38**, 41–52 (1980).
33. Knauer, G. A. & Asper, V. L. (co chairs) *US GOFS Planning Report Number 10, Sediment Trap Technology and Sampling* (US GOFS Planning Office, Woods Hole, 1989).

Substantial hydrogen solubility in olivine and implications for water storage in the mantle

Quan Bai & D. L. Kohlstedt

Department of Geology and Geophysics, University of Minnesota,
310 Pillsbury Drive S.E., Minneapolis, Minnesota 55455, USA

WATER plays an important role in geodynamic processes in the Earth's upper mantle: for example, hydrogen (as water) affects the amount and composition of magma generated by partial melting¹ and a trace amount of hydrogen (~0.001 wt % water) can markedly weaken the dominant upper-mantle mineral, olivine^{2,3}. Migration of hydrogen ions may be responsible for the anomalously high electrical conductivity of the asthenosphere⁴. The quantitative importance of hydrogen in mantle processes must depend on how much water or hydrogen can be stored in nominally anhydrous olivine, and on where hydrogen resides in the olivine lattice. Here we report the results of hydrothermal experiments on olivine single crystals, which show that at 1,573 K and 50–300 MPa, olivine can accommodate as much as 0.0034 wt % water. Hydrogen solubility depends on hydrogen fugacity and oxygen fugacity to the first and the one-half powers, respectively, indicating that hydrogen ions are associated with either oxygen interstitials or magnesium vacancies. Extrapolation of our data to a depth of ~100 km under oceanic areas yields a substantial hydrogen solubility (0.03 wt % water), demonstrating that olivine may indeed be a primary sink for hydrogen in the upper mantle.

On the basis of recent experiments, it was proposed that hydrogen diffuses into the olivine lattice as positively charged ions (or protons) and becomes associated with negatively charged point defects such as cation vacancies or anion interstitials^{5,6}. The solubility of hydrogen in olivine is thus expected to be coupled to the population of these point defects. Therefore for a quasi-ternary silicate such as olivine in an aqueous environment, the Gibbs phase rule⁷ requires that the solubility of hydrogen depends on temperature (T), pressure (P), oxygen fugacity (f_{O_2}), orthopyroxene activity (a_{opx}) and hydrogen fugacity (f_{H_2}). This paper reports the results of an experimental investigation of (1) the solubility of hydrogen in the olivine lattice as a function of f_{O_2} and f_{H_2} at $a_{opx} = 1$ and $T = 1,573$ K, and (2) the type of point defects with which hydrogen becomes associated. (In the upper mantle, a_{opx} is unity because olivine coexists with orthopyroxene, and $T = 1,573$ K corresponds to the temperature at a depth of ~100 km beneath the ocean.)

Samples were prepared from single crystals of olivine from San Carlos, Arizona, of approximate composition $(Mg_{0.9}Fe_{0.9})_2SiO_4$. Four to five samples were fitted into square holes in talc discs and, with a few milligrams of distilled water, were then enclosed in either a nickel or an iron capsule which was subsequently welded shut. Next, the samples were hydrothermally annealed in a gas-medium high-pressure, high-temperature apparatus. The f_{O_2} was controlled at $10^{-6.3}$ or $10^{-10.5}$ atm., respectively, by the partial oxidation of Ni or Fe. As a water-containing fluid phase was found in the capsules at the end of the hydrothermal anneals, the hydrogen fugacity f_{H_2} ,

Received 21 October 1991; accepted 12 May 1992.

1. Billétt, D. S. M., Lampitt, R. S., Rice, A. L. & Mantoura, R. F. C. *Nature* **302**, 520–522 (1983).
2. Alldredge, A. L. & Gotschalk, C. *Limnol. Oceanogr.* **33**, 339–351 (1988).
3. Asper, V. L. *Deep Sea Res.* **34**, 1–17 (1987).
4. Urrere, M. A. & Knauer, G. A. *J. Plank. Res.* **3**, 369–387 (1981).
5. Deuser, W. G. & Ross, E. H. *Nature* **283**, 364–365 (1980).
6. Deuser, W. G. *Deep Sea Res.* **33**, 225–246 (1986).
7. Lohrenz, S. E. *et al. Deep Sea Res.* (in press).
8. Karl, D. M., Knauer, G. A. & Martin, J. H. *Nature* **332**, 438–441 (1988).
9. Michaels, A. F., Silver, M. W., Gowing, M. M. & Knauer, G. A. *Deep Sea Res.* **37**, 1285–1296 (1990).
10. Gardner, W. D., Southard, J. B. & Hollister, C. D. *Mar. Geol.* **65**, 199–242 (1985).
11. Honjo, S., Spencer, D. W. & Farrington, J. W. *Science* **216**, 516–518 (1982).
12. Knauer, G. A. & Martin, J. H. *Limnol. Oceanogr.* **26**, 181–186 (1981).
13. Fitzwater, S. E., Knauer, G. A. & Martin, J. H. *Limnol. Oceanogr.* **27**, 544–551 (1982).
14. Knauer, G. A., Martin, J. H. & Bruland, K. W. *Deep Sea Res.* **26**, 97–108 (1979).
15. Honjo, S. & Doherty, K. W. *Deep Sea Res.* **35**, 133–149 (1988).
16. Honjo, S. *J. Mar. Res.* **36**, 469–492 (1978).
17. Deuser, W. G., Ross, E. H. & Anderson, R. F. *Deep Sea Res.* **28**, 495–505 (1981).
18. Jickells, T. D., Church, T. M. & Deuser, W. G. *Global biogeochem. Cycles* **1**, 117–130 (1987).
19. Jickells, T. D., Deuser, W. G. & Belastock, R. A. *Mar. Chem.* **29**, 203–219 (1990).
20. Honjo, S. *Science* **218**, 883–884 (1982).
21. Izdar, E. *et al. Naturwissenschaften* **71**, 478–479 (1984).
22. Deuser, W. G. *Deep Sea Res.* **37**, 1331–1343 (1990).
23. Walsh, I., Fisher, K., Murray, D. & Dymond, J. *Deep Sea Res.* **35**, 59–70 (1988).
24. Honjo, S., Manganini, S. J. & Cole, J. J. *Deep Sea Res.* **29**, 609–625 (1982).
25. Honjo, S. *J. Mar. Res.* **38**, 53–97 (1980).

as well as the water fugacity $f_{\text{H}_2\text{O}}$, was calculated⁵ on the basis of equilibrium thermodynamics using the fugacity coefficients for H_2O and H_2 (refs 8, 9), the reaction constant for dissociation of water¹⁰, and the values of f_{O_2} and P . For experiments using Ni capsules, $f_{\text{H}_2} = 10^{0.5}$, $10^{1.0}$ and $10^{1.4}$ atm for $P = 50$, 150 and 300 MPa, respectively; for experiments using Fe capsules, $f_{\text{H}_2} = 10^{2.3}$, $10^{2.9}$ and $10^{3.2}$ atm for $P = 50$, 150 and 300 MPa, respectively. During the experiments, talc decomposed into water (which supplied additional hydrogen) and enstatite (which controlled a_{opx} at unity). On the basis of published values for hydrogen diffusivity³, the samples were saturated with hydrogen after annealing for 4 to 6 h at 1,573 K. Under these conditions, cation and anion vacancies and interstitials as well as the hydrogen ions were all sufficiently mobile that their concentrations were set by the imposed T , P , f_{O_2} , a_{opx} and f_{H_2} conditions^{5,11}. The hydrogen concentration in the hydrothermally annealed samples was determined by analysis of the O–H stretching bands in Fourier transform infrared spectra¹², obtained with the infrared beam directed along the [010] direction. The analysis procedure described in ref. 12 may overestimate the hydrogen concentration¹³.

Typical infrared spectra for hydrothermally annealed samples (Fig. 1) reveal that the heights of the infrared bands increase with increasing f_{H_2} at constant f_{O_2} and that the location of the infrared bands are the same at both f_{O_2} values (Ni/NiO and Fe/FeO). The latter point indicates that the sites into which hydrogen becomes incorporated are insensitive to a change in f_{O_2} . The concentration of hydrogen was calculated separately for bands at wavenumbers in the range 3,450–3,800 cm^{-1} (group I bands) and for those in the range 3,000–3,450 cm^{-1} (group II bands), because the two groups of bands are probably associated with different point defects⁶. Here only the group I bands are considered, because the concentration of hydrogen associated with the group II bands is about one order of magnitude smaller than that associated with the group I bands. Hence, in this respect, the values reported here provide a lower limit to the solubility of hydrogen in olivine, especially at high oxygen fugacities⁶.

The solubility of hydrogen (C_{H}), plotted as a function of f_{H_2} in Fig. 2a, was analysed according to the power law relation

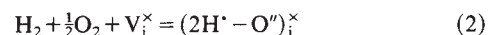
obtained from point-defect chemistry⁶

$$C_{\text{H}} = A f_{\text{O}_2}^m f_{\text{H}_2}^s \quad (1)$$

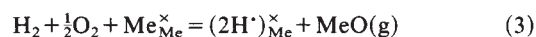
where m and s are constants and A is a function of T , P and a_{opx} . Within experimental scatter, the f_{H_2} exponent $s \approx 1$ for group I bands is about the same at both f_{O_2} values, with $s \approx 0.8$ and 1.0 for the Ni/NiO and Fe/FeO buffers, respectively. This result further supports the conclusion that the sites into which hydrogen becomes incorporated are independent of f_{O_2} . Hence, an f_{O_2} exponent $m = 0.5$ could be calculated from the difference in hydrogen solubility between the data sets at the two f_{O_2} values; the factor $A = 27,542$ for f_{O_2} and f_{H_2} in atmospheres and C_{H} in $\text{H}/10^6\text{Si}$.

The solubility of hydrogen, plotted against $f_{\text{H}_2\text{O}}$ in Fig. 2b, was also analysed according to a power law relation, $C_{\text{H}} \propto f_{\text{H}_2\text{O}}^w$, because $f_{\text{H}_2\text{O}} \propto f_{\text{O}_2}^{1/2} f_{\text{H}_2}$. A least-squares fit of these data yields $C_{\text{H}} = 0.294 f_{\text{H}_2\text{O}}^{0.8}$, with $f_{\text{H}_2\text{O}}$ in atm. It should be noted at this point that, because the activation volume for the formation of hydrogen defects is of the order of the volume of a hydrogen ion, the effect of total pressure on hydrogen solubility will lower C_{H} by less than a factor of two even at $P = 10$ GPa. Thus, the effect of P on C_{H} through an activation volume term can be ignored.

The experimental results bear the following geophysical implications: first, two models for the incorporation mechanism of hydrogen associated with group I bands are proposed based on the measured values of $s \approx 1$ and $m \approx 1/2$ combined with point-defect thermodynamics. In the first model, hydrogen is incorporated with doubly negatively charged oxygen interstitials O_i^{2-} (the superscript indicates the charge and the subscript the site¹⁴):



where $\text{V}_i^{\times}$ denotes a neutral vacant interstitial site and $(\text{2H}^+ - \text{O}^{2-})_i^{\times}$ denotes association of two positively charged hydrogen ions, H^+ , with one O_i^{2-} . From the law of mass action, the concentration of $(\text{2H}^+ - \text{O}^{2-})_i^{\times}$ is proportional to $f_{\text{O}_2}^{1/2} f_{\text{H}_2}$, which in turn is proportional to $f_{\text{H}_2\text{O}}$, in good agreement with our experimental results. In the second model, hydrogen is associated with doubly negatively charged metal (using Me to represent Mg, Fe) vacancies $\text{V}_{\text{Me}}^{2-}$:



where $(\text{2H}^+)_i^{\times}$ denotes two H^+ incorporated into one vacant metal site and $\text{MeO}(\text{g})$ indicates a metal oxide molecule released to the gas phase. For this mechanism, the concentration of

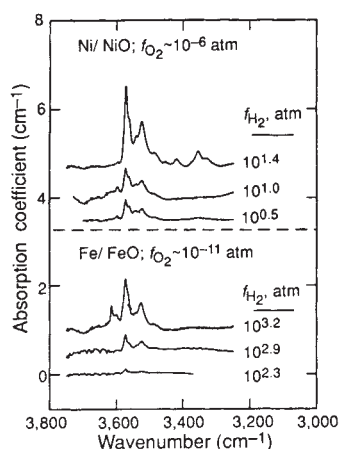


FIG. 1 Typical infrared absorption spectra for hydrothermally annealed samples, obtained with the infrared beam parallel to the [010] crystallographic direction of olivine. The three spectra in the top portion are for an f_{O_2} set by a Ni/NiO buffer; the three in the lower portion are for an f_{O_2} set by an Fe/FeO buffer. Hydrogen fugacity was varied by changing total pressure for a given oxygen fugacity. The top spectrum contains sharp peaks (group II) at wavenumbers between 3,200 and 3,450 cm^{-1} , but these peaks merge into a broad band in the other spectra. To facilitate comparison of the infrared data from different samples, the spectra have been shifted along the vertical axis.

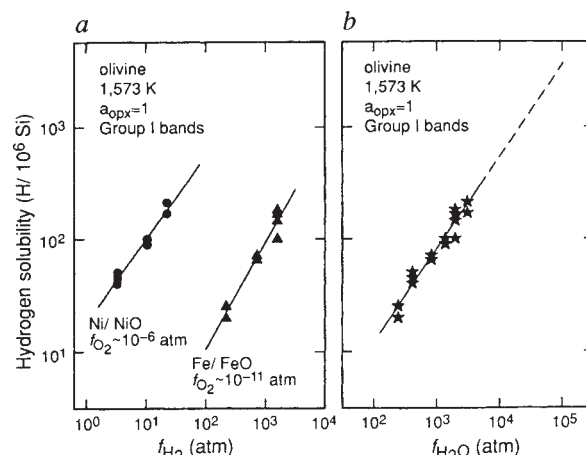


FIG. 2 a, Plot of hydrogen solubility versus hydrogen fugacity. b, Plot of hydrogen solubility versus water fugacity.

$(2H^+)_{Me}$ is also proportional to $f_{O_2}^{1/2}f_{H_2}$, again in good agreement with the measured values.

Second, the experimentally determined relation (Fig. 2) between hydrogen solubility and water fugacity (or hydrogen fugacity) can be extrapolated to upper mantle conditions at 1,573 K if the value of f_{H_2O} (or f_{O_2} and f_{H_2}) is known. At a depth of ~100 km under an oceanic area ($P \approx 3$ GPa and $T \approx 1,573$ K; refs 15, 16), amphibole breaks down into opx, garnet and an aqueous fluid phase¹⁷. If H_2O , O_2 and H_2 are the major components of the fluid phase, the water fugacity will be $f_{H_2O} \approx 2 \times 10^5$ atm, based on published values for the fugacity coefficients for H_2O and H_2 (refs 8, 9), the reaction constant for dissociation of water¹⁰, and the range of f_{O_2} expected to occur in the upper mantle^{18–20}. To calculate the water fugacity, the water fugacity coefficient at 3 GPa was obtained from a polynomial extrapolation of published values, which lie in the range 0.1 to 1 GPa (ref. 8). This extrapolation significantly underestimates the single published value for the water fugacity at 10 GPa (ref. 8). Hence, the calculated value for f_{H_2O} is probably a lower limit. From Fig. 2b, the solubility of hydrogen in olivine at $f_{H_2O} \approx 2 \times 10^5$ atm is 5,100 H/10⁶Si or 0.03 wt % H_2O . As the water concentration in the upper mantle is estimated to be in the range 0.01–0.1 wt % (refs 21, 22), grains of olivine could be a principal storage site for water in this region of the Earth. Hence hydrous phases (such as phases A, B and C (ref. 23)) may not be necessary to account for the water content of the upper mantle.

Finally, it should be pointed out that our extrapolated value for the water solubility level is at least as high as that reported for mantle-derived orthopyroxene grains, but about a factor of ten larger than that in olivine samples from mantle xenoliths¹³.

This discrepancy may develop because hydrogen can diffuse rapidly out of olivine during transport of xenoliths to the Earth's surface⁵, whereas it is possible that hydrogen cannot be readily removed from orthopyroxene grains (ref. 24; and S. J. Mackwell, personal communication). With its substantial water solubility, olivine—the most abundant mineral in the Earth's upper mantle—may dominate the water budget in the mantle. □

Received 27 January; accepted 12 May 1992.

1. Wilson, M. *Igneous Petrogenesis: A Global Tectonic Approach* 54 (Hyman, London, 1989).
2. Mackwell, S. J., Kohlstedt, D. L. & Paterson, M. S. *J. geophys. Res.* **90**, 11319–11333 (1985).
3. Karato, S., Paterson, M. S. & Fitz Gerald, J. D. *J. geophys. Res.* **91**, 8151–8176 (1986).
4. Karato, S. *Nature* **347**, 272–273 (1990).
5. Mackwell, S. J. & Kohlstedt, D. L. *J. geophys. Res.* **95**, 5079–5088 (1990).
6. Bai, Qian & Kohlstedt, D. L. *Phys. Chem. Miner.* (submitted).
7. Willard Gibbs, J. *The Scientific Papers of J. Willard Gibbs* (Dover, New York, 1961).
8. Tödeheide, K. in *Water: A Comprehensive Treatise 1: The Physical Chemistry of Water* (ed. Franks, F.) 463–514 (Plenum, New York, 1972).
9. Shaw, H. R. & Wones, D. R. *Am. J. Sci.* **262**, 918–929 (1964).
10. Robie, R. A. *et al. U.S. geol. Surv. Bull.* **1452**, 172 (1978).
11. Mackwell, S. J., Dimos, D. & Kohlstedt, D. L. *Phil. Mag.* **A57**, 779–789 (1988).
12. Paterson, M. S. *Bull. Miner.* **105**, 20–29 (1982).
13. Bell, D. R. & Rossman, G. R. *Science* **255**, 1391–1397 (1992).
14. Kröger, F. A. & Virk, H. J. *Solid State Physics Vol. 3* (ed. Seitz F. & Turnbull, D.) 307–435 (Academic, New York, 1956).
15. Mercier, J. C. & Carter, N. L. *J. geophys. Res.* **80**, 3349–3362 (1975).
16. Mercier, J. C. *Tectonophysics* **70**, 1–37 (1980).
17. Lambert, T. B. & Wyllie, P. J. *Nature* **219**, 1240–1241 (1968).
18. Eggler, D. H. *Geophys. Res. Lett.* **10**, 365–368 (1983).
19. Wood, B. J., Bryndzia, L. T. & Johnson, K. E. *Science* **248**, 337–345 (1990).
20. Ballhaus, C., Berry, R. F. & Green, D. H. *Nature* **348**, 437–440 (1990).
21. Ringwood, A. E. *Composition and Petrology of the Earth's Upper Mantle* (McGraw-Hill, New York, 1975).
22. Michael, P. J. *Geochim. cosmochim. Acta* **52**, 555–566 (1988).
23. Ringwood, A. E. & Major, A. *Earth planet. Sci. Lett.* **2**, 130–133 (1967).
24. Skogby, H. & Rossman, G. R. *Am. mineral.* **74**, 1059–1069 (1989).

ACKNOWLEDGEMENTS. We thank S. Mackwell, S. Karato, G. Hirth and Z.-M. Jin for discussions. This research was supported by the US NSF.

Grain-boundary graphite in Kapuskasing gneisses and implications for lower-crustal conductivity

Marianne Mareschal*, William S. Fyfe†, John Percival‡ & Tammy Chan†

* Département de Génie Minéral, École Polytechnique, CP 6079, Succursale A, Montréal H3C 3A7, Canada

† Department of Geology, University of Western Ontario, London N6A 5B7, Canada

‡ Geological Survey of Canada, 601 Booth Street, Ottawa K1A 0E8, Canada

SURFACE electromagnetic measurements commonly reveal that the intermediate and lower continental crust has an appreciable electrical conductivity, much higher than that found for laboratory analogues of deep-crustal rocks or for upper-crustal rocks either formed at shallow depths or uplifted from deeper levels. Conducting films at grain boundaries are thought to play an important part in enhancing deep-crustal conductivity, with brines and graphite being the main candidates¹. At present, however, there is a lack of direct evidence for such films. Frost *et al.*² have reported an enhancement in conductivity of igneous lower-crustal rocks from the Laramie complex apparently owing to the presence of graphite films, but the possible role of brines was not clear from this work. Here we present a study of the grain surface composition of metamorphic (and one igneous) rocks exhumed from 20 km depth by the Kapuskasing uplift of the Canadian shield³. We expect these samples to be representative of metamorphic and igneous rocks of the deep Archaean crust. Auger spectroscopy

provides evidence for graphite films at grain boundaries, which can account for the present electrical signature of the Kapuskasing crust^{4–6}. We also detect traces of chlorine, sulphur and iron, which suggest that brines and solid conductors other than graphite may have helped to enhance the conductivity, but their abundance suggests at best a minor role relative to graphite.

The presence of brines at grain boundaries is the explanation favoured by most geophysicists to explain the electrical signature of many deep crusts of Phanerozoic age^{7–9}. Once the fluids had leaked into the upper crust by disrupting an impermeable zone at the ductile–brittle transition level, they would be expelled quickly¹⁰. Various mechanisms have been suggested to explain the presence of this impermeable cap at midcrustal depths, but most are tenable only in relatively young tectonic areas. Elsewhere, they face the problem of fluid retention over timescales much larger than those allowed by pressure and temperature conditions in the deep crust¹¹. This is particularly true of older shields, which seem to be characterized by rather dry lower-crustal material¹² but still have conductivities two orders of magnitude larger than that expected for dry rocks at these depths¹³. Solid conductors such as graphite are therefore being considered as possible alternative sources of conductivity in shield environments.

Crustal cross-sections exposed by differential uplift and erosion hold promise for advancing our knowledge in these areas; this is especially true of the Kapuskasing uplift of the Canadian shield which seems to have survived exhumation from at least 20 km relatively intact³. Although the widespread connectivity of solid films necessary to explain the lower-crustal conductivity would be destroyed by expansion during uplift, traces of conductive material such as sulphides or graphite should still coat grain surfaces and be detectable by high-resolution scanning Auger spectrometry^{2,6}. Similarly, a rock that has transiently hosted pervasive hydrous fluids should contain brine residues at grain boundaries. In a search for these features,

we have systematically scanned six metamorphic rocks and one igneous rock from the Kapuskasing uplift for traces of carbon, iron, sulphur and chlorine. The samples represent meta-tonalites (PBA80-857, PBA80-857B and PBA85-16), paragneisses (PBA79-32 and PBA79-215), anorthosite (PBA80-604) and a low-grade shear zone (PBA88-09). All metamorphic specimens except the last were selected because they correspond to a structural depth of 20–25 km, as measured by geobarometry of the metamorphic assemblages². Specimen PBA88-09 was selected as a control reference, the only sample not representative of the intermediate to lower crust.

Of the seven samples, only the shear zone (PBA88-09) does not contain carbon at grain boundaries. All others do, as shown for instance for sample PBA79-32 in Fig. 1 and for sample PBA85-16 in Fig. 2, although graphite was identified petrographically only in sample PBA79-215 (ref. 14). The Auger spectra also show that calcium is abundant, but the ratios of carbon to calcium are so varied that the carbon response cannot be due to calcium carbonate alone; the calcium is more likely to be a plagioclase component. The carbon maps (Figs 1 and 2, bottom) clearly show that the surface carbon distribution is restricted to grain boundaries, except for rare graphite grains (Fig. 1). The carbon films reach a maximum thickness of ~ 300 Å for the anorthosite sample and can be as thin as 30 Å (sample PBA80-857). Figure 3 indicates that the carbon peak shapes in the Auger spectra of samples PBA79-32 and PBA85-16 are similar to that of pure graphite from Sri Lanka. In other samples or at other scanning locations, the presence of calcium or potassium, which have Auger peaks adjacent to that of carbon, has obscured the information. We obtained laser Raman spectra, produced by a microprobe with sample spot sizes of ~ 1 μm , for some samples; they show evidence for disordered graphitic carbon as would be produced by ultra-fine films ~ 50 Å thick.

Iron is also abundant in the Auger spectra but seems to correspond to opaque grains previously recognized from petrography¹⁴. Traces of chlorine and sulphur are sporadically present: of the seven samples analysed, only two contain much chlorine. Even for these specimens, the Auger spectra show that chlorine is much less abundant than carbon: a maximum of 1% chlorine or sulphur is seen, compared with typical 5% carbon values at any scanning location (Fig. 4 for sample PBA80-16). The evidence is therefore weak for transient passage of brines through these rocks, so they seem unlikely to be the conductive source in the deep Archean crust. There is strong evidence that if high chlorine content is once present in a rock, chlorine is preserved on grain boundaries. For instance, work in progress on shear zones in the Brazilian iron formations of Minas Gerais also shows large amounts of graphite but in association with appreciable chlorine contents (5–7% each), indicating that these rocks hosted the passage, possibly sequential, of CO_2 plus CO , saline waters and/or hydrocarbon-rich fluids. Our results suggest that the Kapuskasing rocks formed or cooled under conditions that were favourable to graphite precipitation, whereas evidence of hydrous, saline fluids is sparse. This conclusion is supported by oxygen isotope studies of Kapuskasing samples whose heterogeneity suggests that no pervasive fluid phase was involved during high-grade regional metamorphism¹⁵.

Clearly, our results cannot directly constrain the sources of electrical conductivity in the present deep levels of the Canadian shield. It is not yet possible to relate unambiguously observations on rocks formed more than 2×10^9 years ago at mid- to lower-crustal levels to present conditions at these depths. A low-resolution electromagnetic survey¹⁶ originally showed that no overall electrical anomaly was associated with the Kapuskasing uplift, confirming that rocks formed at great depth had lost any conductivity they might have had during exhumation. But

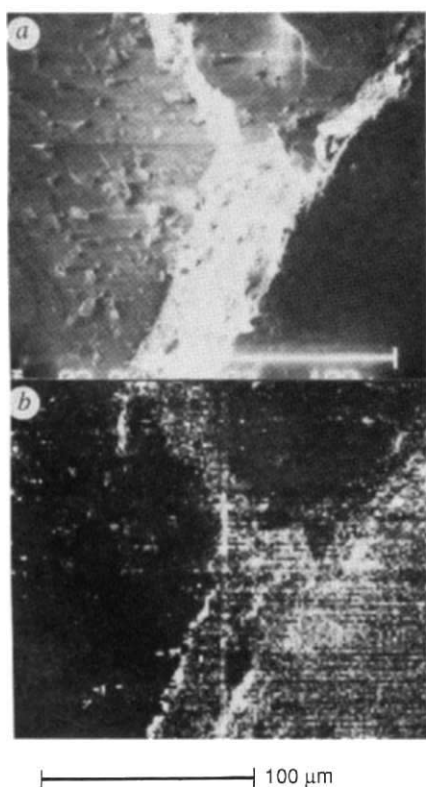


FIG. 1 a, Scanning electron micrograph for sample PBA79-32; b, Auger spectrometry carbon map for same area. White represents carbon-rich zones.

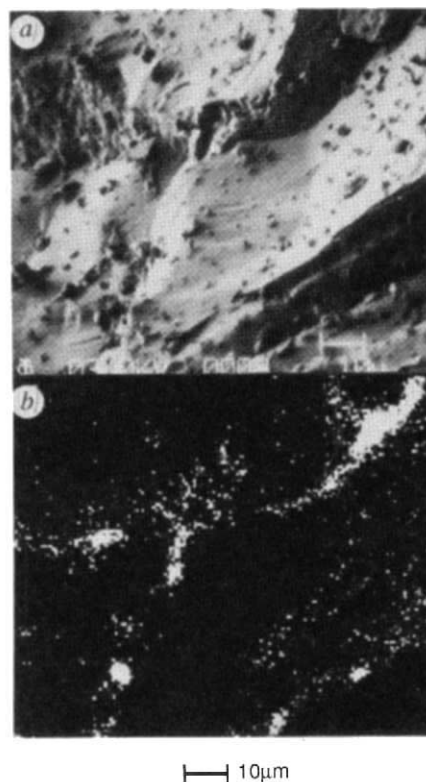


FIG. 2 a, Scanning electron micrograph for sample PBA85-16; b, Auger spectrometry carbon map for same area. Note the discontinuity of carbon films at grain boundaries.

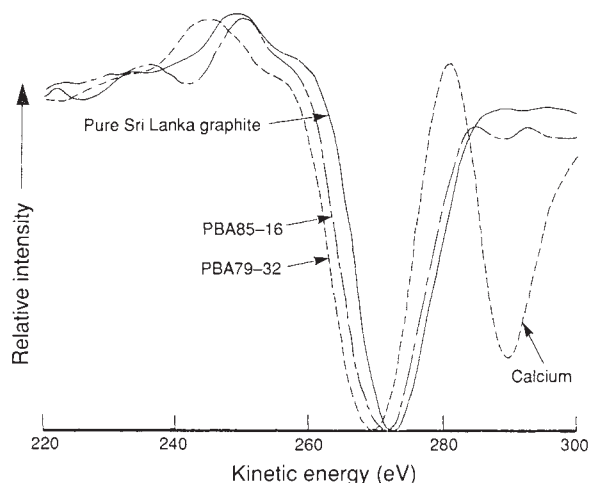


FIG. 3 Comparison of Auger spectra from grain-boundary graphite in samples PBA79-32 and PBA85-16 with a spectrum from pure Sri Lanka graphite.

detailed higher-resolution surveys have since revealed the presence of discrete conductors, with typical resistivities of $10^4 \Omega \text{ m}$ against a $10^5 \Omega \text{ m}$ background, in the present upper crust of the uplifted zone¹⁷⁻¹⁹, as well as the presence of electrical anisotropy in the lower crust and/or upper mantle^{20,21}. Deep anisotropy is a good indicator of solid conduction in the shield because fluids would tend to re-equilibrate or escape very quickly. For instance, heat-flow observations over the Canadian shield suggest temperatures of $\sim 300^\circ \text{C}$ at the Moho²², implying that a predominantly mafic intermediate to lower crust could be entirely brittle, a hypothesis supported by the deep earthquakes (depth $> 25 \text{ km}$) recorded in the region²³. No mechanism could trap mechanically stable free water for long under these conditions¹¹. On the other hand, some of the shallow upper-crustal conductive zones, particularly those that are subhorizontal and cut through known structural features, must post-date the uplift and probably image the pervasive brines commonly found in shields^{17,24}. Other zones seem to relate to dipping seismic reflectors, possibly imaging a laminated gneissic zone of deep-crustal origin now uplifted to the top few kilometres of the crust¹⁹, and may be more representative of original crustal properties. The rock samples that we analysed come from very near this detachment zone; therefore, we measured in the laboratory both the complex electrical conductivity of sample PBA79-32 (whose Auger spectrometry carbon map is shown in Fig. 1) and its variation with frequency from 10^6 to 1 Hz (at room temperature), a procedure which allows one to distinguish between surface and bulk conductivity⁴.

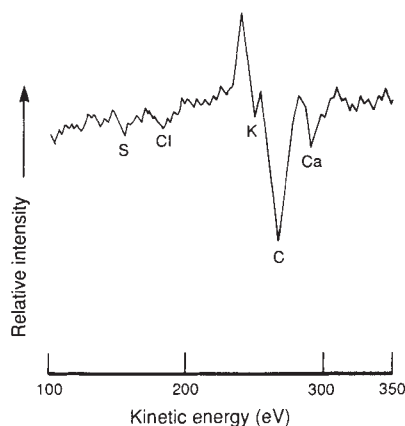


FIG. 4 Section of Auger spectrum for sample PBA85-16 showing the relative importance of sulphur, chlorine, carbon and iron in a chlorine-rich sample.

Model calculations⁶ based on these laboratory measurements indicate that the strong surface conduction effects observed in the sample can be explained by the presence of fragmented graphite films at grain boundaries. The model also shows that bulk resistivities of a few $100 \Omega \text{ m}$, as typically observed in the lower crust of shields, could be produced by pore lining with thin connected graphite films ($50\text{--}200 \text{ \AA}$); such graphitic rocks, once uplifted to the surface, could still display resistivities 10 times smaller (a few times $10^3 \Omega \text{ m}^{-1}$ as observed on sample PBA79-32) than those measured on similar rocks devoid of graphite (several times $10^4 \Omega \text{ m}$) even if the films are then disconnected. Graphite grain-boundary films (or what remains of the coating once the rock is uplifted to the surface) could therefore explain equally well the lower-crustal conductivities observed in the Kapuskasing structural zone and some of the upper-crustal features slightly less resistive than the background. The electrical measurements on sample PBA79-32, as well as work in process on the other specimens⁵, further indicates that the conductivity along the foliation is higher by a factor of 2 to 3 than that across the foliation, suggesting that the present electrical signature of the deep Canadian shield may be partly explained by a pervasive micro-anisotropy related to foliation. This anisotropy could be enhanced by the presence of localized conductive deposits in veins or fractures, by resistive magmatic bodies emplaced into orientations controlled by the prevailing stress field, or by other features more directly related to mineralogical assemblages. $\square$

Received 6 December 1991; accepted 29 April 1992.

1. Jones, A. G. in *Continental Lower Crusts* (eds Fountain, D. M., Arculus, R. J. & Kay, R. W.) (Elsevier, New York, in the press).
2. Frost, B. R., Fyfe, W. S., Tazaki, K. & Chan, T. *Nature* **340**, 134–136 (1989).
3. Percival, J. A. in *Reflection Seismology: the Continental Crust* (eds Barazangi, M. & Brown, L.) 135–142 (Am. geophys. Un., Washington DC, 1986).
4. Katsube, T. J., Mareschal, M. & Aucoin, F. *Geol. Survey Can. Current Res. Pap.* 91-E, 257–263 (1991).
5. Katsube, T. J., Scromeda, N., Mareschal, M. & Bailey, R. C. *Geol. Survey Can. Current Res.* (in the press).
6. Katsube, T. J. & Mareschal, M. *J. geophys. Res.* (in the press).
7. Gough, D. I. *Nature* **323**, 143–144 (1986).
8. Jones, A. G. *Geophys. J. Int.* **89**, 7–18 (1987).
9. Hyndman, R. D. & Shearer, P. M. *Geophys. J. Int.* **98**, 343–365 (1989).
10. McCaig, A. M. *Geology* **9**, 867–870 (1988).
11. Bailey, R. C. *Geophys. Res. Lett.* **17**, 1129–1132 (1990).
12. Yardley, B. W. D. *Nature* **323**, 111 (1986).
13. Haak, V. & Hutton, R. S. in *The Nature of the Lower Continental Crust* (eds Dawson, J. B., Carswell, D. A., Hall, J. & Wedepohl, K. H.) 35–49 (Geol. Soc. London, 1986).
14. Percival, J. A. *Am. Mineral.* **68**, 667–686 (1983).
15. Li, H., Schwarcz, H. P. & Shaw, D. M. *Contrib. Mineral. Petrol.* **107**, 448–458 (1991).
16. Woods, D. V. & Allard, M. *Phys. Earth planet. Inter.* **42**, 135–142 (1986).
17. Bailey, R. C., Craven, J. A., Macnae, J. C. & Polzer, B. D. *Nature* **340**, 136–138 (1989).
18. Kurtz, R. D., Macnae, J. C. & West, G. F. *Geophys. J. Int.* **99**, 195–203 (1989).
19. Percival, J. A. *et al. EOS* **32**, 337–341 (1991).
20. Kurtz, R. D., Craven, J. A., Niblett, E. R. & Stevens, R. A. *Geophys. J. Int.* (submitted).
21. Kellett, R. L., Mareschal, M. & Kurtz, R. D. *Geophys. J. Int.* (in the press).
22. Mareschal, J. C. *Tectonophysics* **194**, 349–356 (1991).
23. North, R. G. *et al. Seism. Res. Lett.* **60**, 89–93 (1989).
24. Frape, S. K. & Fritz, P. in *Saline Water and Gases in Crystalline Rocks* (eds Fritz, P. & Frape, S. K.) 19–38 (Geol. Assoc. Canada, 1987).

Ages of altered granites adjoining the Witwatersrand Basin with implications for the origin of gold and uranium

L. J. Robb*, D. W. Davis†, S. L. Kamo† & F. M. Meyer*

* Department of Geology, University of the Witwatersrand, 1 Jan Smuts Avenue, Johannesburg 2050, Republic of South Africa

† Jack Satterley Geochronology Laboratory, Royal Ontario Museum, Toronto, Ontario M5S 2C6, Canada

ONE of the main problems in understanding the origin of the Witwatersrand Basin mineralization concerns the source of the prodigious concentrations of gold and uranium in the basin; another problem is the age and nature of the rocks in the source area. Recent data^{1,2} indicate that the source area was largely 3,200–2,900 Myr old, becoming progressively younger as the basin evolved. Granites adjoining the basin are commonly overprinted by pervasive and vein-related hydrothermal alteration, which resulted in an enrichment of both gold and uranium^{3,4}. Here we report uranium–lead data demonstrating that these hydrothermally altered granites were intruded sporadically, both before and during the 360-Myr period⁵ constraining Witwatersrand sediment deposition, and that hydrothermal alteration was associated with granitoid intrusion and subsequent cooling. Thus, ‘fertilization’ of the crust occurred during basin evolution, a feature that

supports the idea¹⁹ that gold and uranium mineralization occurred simultaneously with sediment deposition.

The Witwatersrand Basin has produced nearly 40% of all the world's gold and still contains reserves of ~40,000 tonnes (ref. 7). In the late 1970s it also provided the western world with ~20% of its uranium supplies⁸. Conglomerates (extremely coarse sediments) host particulate gold and uraninite (UO₂), which occur along with concentrations of pyrite and other heavy minerals. Despite more than 100 years of exploitation, the origin of the Au–U mineralization in the Witwatersrand Basin remains controversial, and the age of sediment deposition has, until recently, also been poorly constrained.

Mineralization, which is concentrated almost entirely within the upper part of the sequence (the Central Rand group; Fig. 1), is regarded either as the result of concentration of heavy detrital minerals during sedimentation with subsequent modification of allogenic constituents during diagenesis and metamorphism (the modified placer theory⁶), or as an epigenetic concentration of hydrothermal ores after basin formation⁹. The Witwatersrand Basin was originally believed to have been early Proterozoic (2,800–2,300 Myr; ref. 10) in age, but more-precise age constraints indicate that it was deposited episodically over 360 Myr, between 3,074 ± 6 and 2,714 ± 8 Myr ago⁵. A tentative age of 2,914 ± 8 Myr (ref. 5) is also available for the Crown lava, a thin persistent volcanic unit that occurs midway up the sequence (Fig. 1).

We sampled eight localities of hydrothermally altered granites (HAGs) around the edges of the Witwatersrand Basin (Fig. 1). Uranium–lead isotope ratios of zircon, and in certain cases monazite, fluorite and rutile, have been analysed following

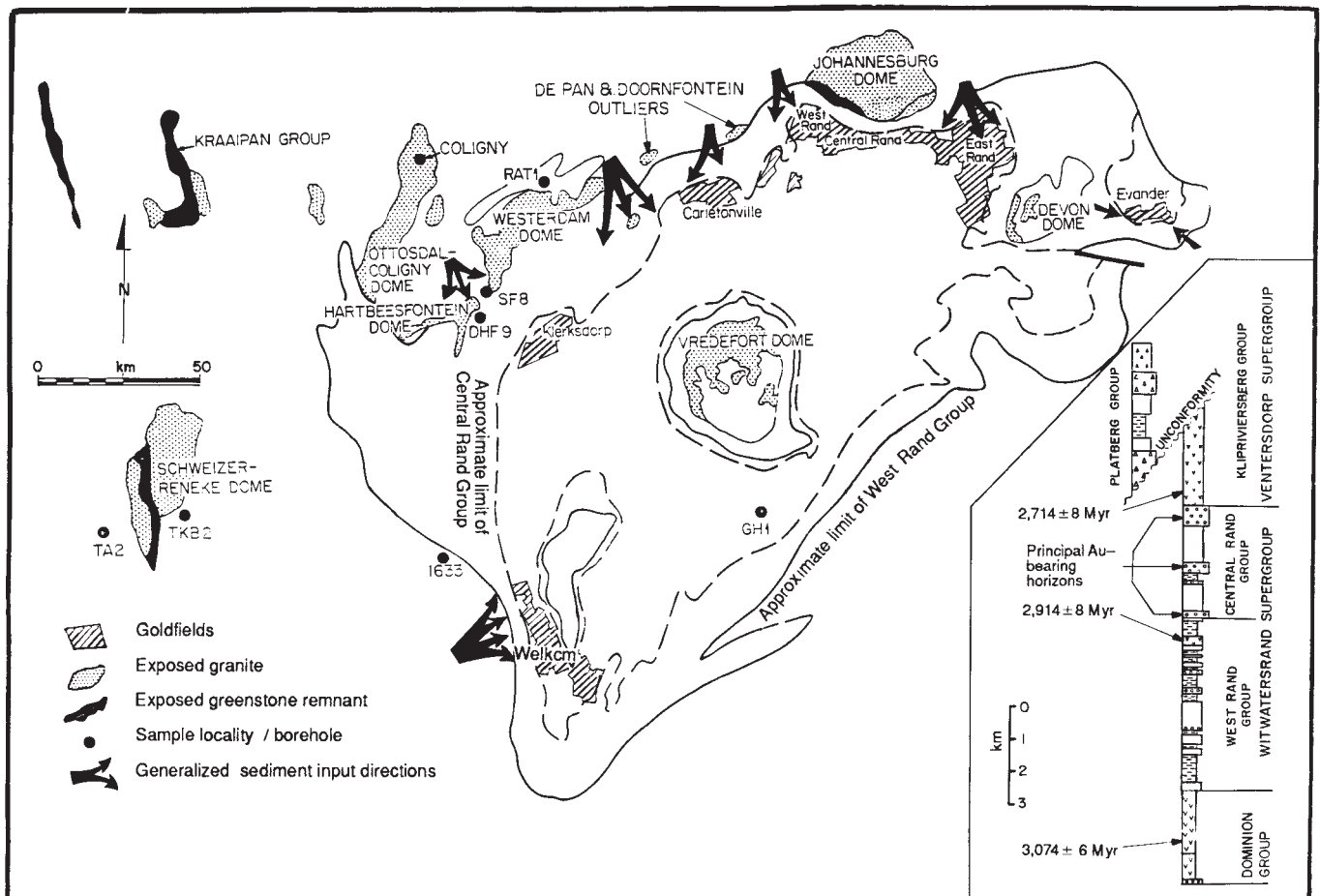


FIG. 1 Schematic plan of the Witwatersrand Basin showing localities of the samples described in the text. Inset shows a simplified stratigraphic column

and key volcanic horizons providing chronostratigraphic control (from recent U–Pb zircon age determinations⁹).

TABLE 1 Characteristics of HAGs adjoining the Witwatersrand Basin

Name locality	Composition	Alteration (vein paragenesis)	Au (p.p.b.)*	U (p.p.m.)	Age (Myr)	Overlain by
Hartbeesfontein DHF9	Peraluminous granodiorite-adamellite (primary garnet)	Greisen (quartz-chlorite-pyrite-chalcopyrite-molybdenite) $\pm$ uraninite	0.7-4.1	2-10	$3,174_{-7}^{+9}$	Dominion
GH1	Low-Ca granite (monazite)	Deuteric (abundant uraninite)	0.7-1.7	6-116	$3,101 \pm 2$	West Rand
Westerdam SF8/RAT1	Granite	(Quartz-chlorite-pyrite-fluorite-carbonate)	0.6-7.4	1-21	$3,086 \pm 3$	West Rand, Dominion
Coligny TA2	Low-Ca leucogranite Tonalite gneiss	Minor albitization Minor deuteric (contains veins derived from Schweizer-Reneke intrusion)	7.4 —	1-2 —	$3,031_{-10}^{+11}$ $2,927_{-6}^{+23}$ (metamorphic?)	Ventersdorp Ventersdorp
Schweizer-Reneke TKB2	Peraluminous adamellite-granite (monazite)	Deuteric. Quartz-K feldspar-chlorite-rutile-pyrite-chalcopyrite) (exogeneous)	6.0-127	1-34	$2,880 \pm 2$	Ventersdorp
1633	Granodiorite-adamellite	(Quartz-chlorite-pyrite-chalcopyrite-galena)	1.3-9.6	3-4	$2,727_{-5}^{+6}$	Mid-Ventersdorp

* Parts per 10^9 .

standard procedures used at the Royal Ontario Museum¹¹⁻¹³. Details of the granite bodies analysed, their compositions and the nature of hydrothermal alteration affecting them are summarized in Table 1. Isotope data are presented in Table 2 and also plotted on concordia diagrams in Fig. 2. We determined concordia intersection ages and errors (quoted at 95% confidence) using published correlation procedures¹⁴.

Three of the granites analysed are older than the Dominion group ($>3,074$ Myr) and, therefore, represent part of the basement on which the late Archaean volcano-sedimentary basins of the region were deposited. One granite (Coligny) is significantly younger (3,031 Myr) than the Dominion group, but its relationship to the lower Witwatersrand beds is not known. Tonalitic gneisses and migmatites from borehole TA2 give an

age of $2,927_{-6}^{+23}$ Myr which may represent the age of migmatization rather than that of the protolith, as identical rocks further south have yielded tentative ion-microprobe ages of 3,250 Myr for zircon cores and 2,940 Myr for metamorphic overgrowths¹⁵. The Schweizer-Reneke granite (TKB2) is significantly younger (2,880 Myr) than the West Rand group, but it is not known whether it pre-dates, or is coeval with, upper Witwatersrand deposition. Granite intersected in borehole 1633 (2,727 Myr) is only slightly older than the lowermost volcanics of the Ventersdorp supergroup (2,714 Myr; ref. 5) and must have been emplaced very late during, or after, Witwatersrand deposition. It is pertinent to note that Central Rand group sediments along the western edge of the Welkom Goldfield were structurally deformed (and in places overturned) just before Ventersdorp

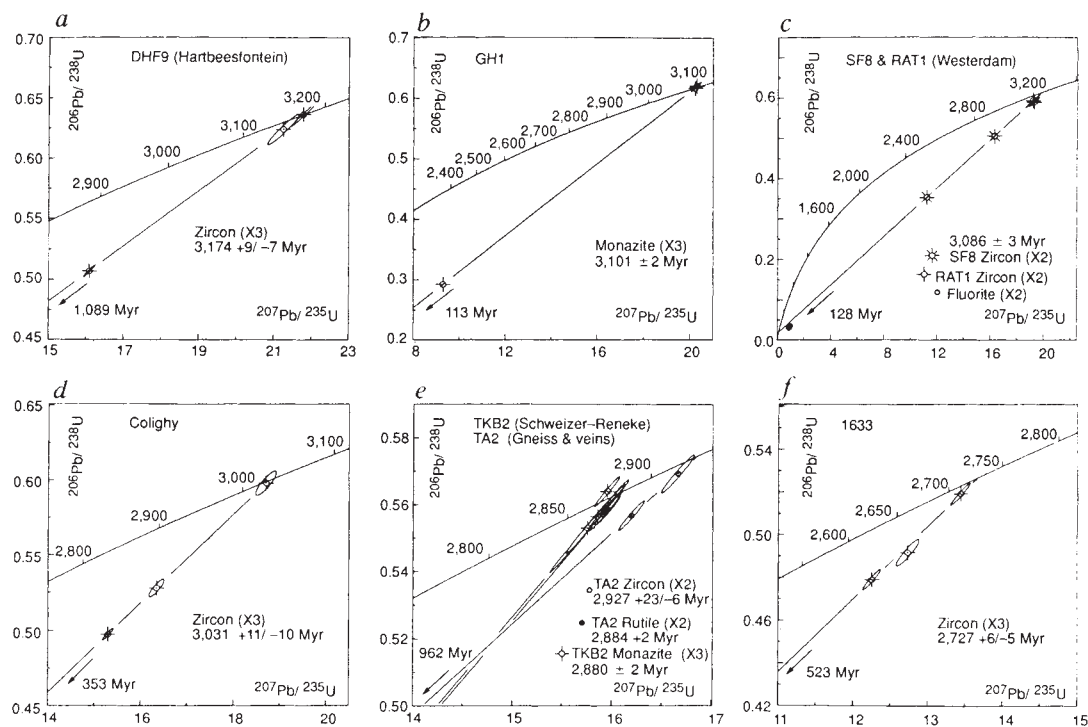


FIG. 2 $^{206}\text{Pb}/^{238}\text{U}$ against $^{207}\text{Pb}/^{235}\text{U}$ concordia plots for granites adjacent to the Witwatersrand Basin. Isotope ratios are presented in Table 2. Figure 2e is a combined plot for data from the TA2 and TKB2 localities. Two zircon fractions from the TA2 tonalite gneisses yield an age of 2,927 Myr. Three

monazite fractions from the TKB2 granite yield an age of 2,880 Myr. Two rutile fractions from a hydrothermal vein cutting the gneissic banding in TA2 provide an age of 2,884 Myr.

TABLE 2 U-Pb isotope data

Sample	Weight (mg)	U (p.p.m.)	Pb _{com} (pg)	Measured ²⁰⁷ Pb/ ²⁰⁴ Pb	Corrected				²⁰⁷ Pb/ ²⁰⁶ Pb age (Myr)
					Th/U	²⁰⁶ Pb/ ²³⁸ U	²⁰⁷ Pb/ ²³⁵ U	% disc.	
DHF9, Hartbeesfontein									
z, Ab, 1 gr	0.003	50	8	219	0.71	0.63653	21.797	0.0 ± 1.6	3,174.2 ± 3.3
z, Ab, 1 gr	0.001	170	3	502	0.65	0.62409	21.256	1.6 ± 2.0	3,165.7 ± 8.9
z, Ab, 3 gr	0.003	80	38	59.7	0.65	0.50677	16.098	16.3 ± 1.0	3,054.6 ± 3.0
r, Ab, 20 gr	0.051	0.3	2,934	20.11	1.64	—	—	−9,000 ± 6	3,323 ± 3
GH1									
M, Ab, 1 gr	0.003	1820	27	1,931	36.9	0.61676	20.198	0.3 ± 0.8	3,103.3 ± 2.6
m, Ab, 1 gr	0.001	3341	88	373	34.5	0.62138	20.286	−0.7 ± 1.0	3,098.3 ± 3.2
m, Ab, 2 gr	0.002	400	8	447	37.6	0.29217	9.306	51.9 ± 1.0	3,059.0 ± 3.2
SF8, Westerdam									
z, Ab, 1 gr	0.002	150	7	334	0.45	0.50561	16.287	17.4 ± 1.0	3,077.0 ± 2.6
z, Ab, 1 gr	0.001	130	7	109.6	0.42	0.35306	11.211	41.7 ± 1.0	3,054.1 ± 3.6
f, Ab, 13 gr	0.028	1.2	6	17.74	0.56	0.02973	0.8816	94.9 ± 4.0	2,944 ± 54
RAT1, Westerdam									
z, Ab, 1 gr	0.005	160	5	1,312	0.60	0.58961	19.286	4.6 ± 2.0	3,101.4 ± 2.7
z, Ab, 2 gr	0.002	130	5	450	0.64	0.59246	19.173	3.4 ± 3.0	3,084.3 ± 3.7
f, Ab, 20 gr	0.209	0.2	11	17.09	2.53	0.03490	0.9646	93.7 ± 8.2	2,830 ± 52
Coligny									
z, Ab, 1 gr	0.001	20	4	65.4	1.26	0.59806	18.703	0.3 ± 1.2	3,029.6 ± 10.8
z, Ab, 1 gr	0.001	30	2	134.7	1.50	0.52794	16.347	11.4 ± 1.0	3,013.6 ± 8.5
z, Ab, 3 gr	0.003	40	4	226.3	1.24	0.49718	15.301	16.2 ± 0.8	3,004.0 ± 3.5
TA2 (gneiss plus vein)									
z, Ab, 2 gr	0.003	130	3	1,077	0.59	0.56925	16.661	0.8 ± 1.0	2,923.0 ± 3.3
z, Ab, 2 gr	0.003	170	6	597	0.68	0.55666	16.197	2.6 ± 0.8	2,913.3 ± 2.6
r, Ab, 1 gr	0.604	7	126	256.4	0.01	0.55850	15.961	1.0 ± 0.8	2,884.2 ± 3.3
r, Ab, 13 gr	0.307	5	41	305	0.00	0.55762	15.920	1.1 ± 1.0	2,882.6 ± 3.4
TKB2, Schweizer-Reneke									
m, Ab, 1 gr	0.002	9800	22	6,555	13.4	0.55647	15.857	1.2 ± 2.0	2,879.5 ± 3.2
m, Ab, 1 gr	0.001	4200	14	2,137	18.7	0.55293	15.755	1.8 ± 2.4	2,879.4 ± 2.3
m, Ab, 1 gr	0.001	8800	83	797	24.7	0.56399	15.955	−0.7 ± 0.8	2,867.7 ± 2.2
1633									
z, Ab, 1 gr	0.008	50	6	404	0.92	0.51904	13.439	1.3 ± 1.0	2,723.0 ± 3.3
z, Ab, 2 gr	0.002	30	3	151.2	0.83	0.49155	12.734	6.5 ± 1.2	2,723.6 ± 7.5
z, Ab, 1 gr	0.001	150	6	164.8	0.80	0.47889	12.254	8.1 ± 1.0	2,703.3 ± 4.0

Abbreviations: z, zircon; m, monazite; r, rutile; f, fluorite; Ab, abraded; gr, grains; Pb_{com}, Common Pb, including blank. Th/U calculated from ²⁰⁸Pb/²⁰⁶Pb ratio and ²⁰⁷Pb/²⁰⁶Pb age assuming concordance; %disc., per cent discordance relative to ²⁰⁷Pb/²⁰⁶Pb age.

volcanism, suggesting that the 1633 granite either was the cause of this disturbance, or was emplaced along major structures associated with the deformation.

The timing of hydrothermal alteration next to the Witwatersrand Basin is difficult to determine because fluid movements have been complex and long-lived. In several areas, however, U-Pb isotope dating provides an indication that the age of hydrothermal alteration was synchronous with the emplacement of the granites. The Schweizer-Reneke granite (TKB2) intrudes a suite of tonalite gneisses and migmatites (TA2) which are themselves cut by rutile-bearing hydrothermal veins (Table 1). Two fractions of dark, euhedral rutile yield overlapping, concordant, isotope ratios with ²⁰⁷Pb/²⁰⁶Pb ages of 2,884 ± 2 Myr (Fig. 2e). This age is the same, within error, as the Schweizer-Reneke granite in borehole TKB2 (2880 ± 2 Myr; Fig. 2e). The veins cutting tonalitic gneisses and migmatites in borehole TA2 therefore represent an exogenous hydrothermal manifestation of the nearby Schweizer-Reneke granite intrusion.

Additional evidence indicating contemporaneity between hydrothermal alteration and granite emplacement comes from U-Pb isotope analyses of rutile associated with greisens in DHF9. The rutile contains a radiogenic lead component which at present is unsupported by uranium, but which gives a ²⁰⁷Pb/²⁰⁶Pb age of 3,323 Myr assuming an initial Pb isotopic composition on the Stacey and Kramers¹⁶ growth curve. This age depends on the value chosen for initial Pb and could be made to agree with the age of the granite itself (3,174 ± 7 Myr) by the choice of a more evolved initial Pb composition. It is

therefore likely that the rutile formed during granite emplacement and not significantly later. Similarly, isotope analyses of fluorite fractions from the Westerdam granite (Table 1) yield ²⁰⁷Pb/²⁰⁶Pb ages of 2,944 ± 54 Myr (SF8) and 2,830 ± 52 Myr (RAT1) which, although imprecise because of the small quantities of Pb and the high levels of discordance (Fig. 2c), indicate overlap between hydrothermal alteration and the span of granite emplacement. The fluorite data points also plot close to the Pb loss line for zircons from the Westerdam granite.

Although most workers favour a syngenetic origin for the conglomerate-hosted Au-U mineralization, with later remobilization of ore during diagenesis and metamorphism⁶, continuing assertions for an epigenetic origin⁹ are stimulated by the inability of the modified placer model to account for the enormous quantities of Au and U that must have been present in the source area. Our data suggest that fertilization of the Witwatersrand hinterland took place as a consequence of extensive hydrothermal activity caused by the intrusion of numerous granitoid bodies, at least some of which were emplaced during Witwatersrand deposition. Certain of these granites are characterized by differentiated chemical signatures and/or peraluminous compositions, and they tend to be uranium-specific (Table 1). They probably represent an adequate source of detrital uraninite for the Witwatersrand deposits⁴. The HAGs that are present with gold-enriched sulphide-bearing veins do not themselves represent an adequate source of gold, but possibly identify the roots of high-level systems that have largely been eroded away. Large volumes of felsic volcanic rocks are now known to occur in the

hinterland. Examples include the 3,074-Myr-old dacitic to rhyolitic upper Dominion group, and the felsic volcanics and agglomerates of the Kraaipan group, also tentatively dated at ~3,070 Myr (L.J.R., D.W.D., S.L.K. and C. R. Anhaeusser). The Gaborone granite in southeastern Botswana, recently dated at 2,790 Myr (D.W.D., L.J.R., D. F. Grobler and M. C. Jackson, manuscript in preparation), is now known to be genetically related to an extensive province of, now largely eroded, rhyolitic volcanics. This province, previously thought to be Proterozoic in age¹⁷, formed during Central Rand group deposition. It may have been a locus for high-level gold mineralization, as minor silver-base metal veins have been reported from remnants of the rhyolites surrounding the Gaborone granite¹⁸. The erosion of high-level felsic plutonic-volcanic suites, mineralized by associated mesothermal to epithermal fluid systems would seem, in the light of our data, to offer a reasonable solution for the source of Witwatersrand gold.

The complex sequence of events involving uranium specificity, consecutive phases of granitoid plutonism, periods of felsic magma extrusion and hydrothermal alteration may explain the concentrations of Au and U in the Witwatersrand depository by providing a framework within which primary concentrations of metals might have taken place in the source area. The protracted emplacement of granite bodies until late in the depositional history of the basin, and the trend of increasingly younger detrital zircon grains upwards in the stratigraphic succession¹, indicate that the Witwatersrand source area was evolving con-

tinuously during basin formation. This might be the main reason why mineralization is almost entirely concentrated in the Central Rand group, whereas the underlying West Rand group remains largely barren. □

Received 12 December 1991; accepted 30 April 1992.

1. Robb, L. J., Davis, D. W. & Kamo, S. L. *J. Geol.* **98**, 311–328 (1990).
2. Barton, E. S. *et al. Econ. Geol.* **84**, 2012–2019 (1989).
3. Robb, L. J. & Meyer, F. M. *Econ. Geol.* **85**, 511–536 (1990).
4. Robb, L. J., Meyer, F. M., Drennan, G. R. & Ferraz, M. F. S. *Afr. J. Geol.* **93**, 5–40 (1990).
5. Armstrong, R. A., Compston, W., Retief, E. A., Williams, I. S. & Welke, H. J. *Precamb. Res.* **53**, 243–266 (1991).
6. Feather, C. E. & Koen, G. M. *Miner. Sci. Eng.* **7**, 189–224 (1975).
7. Handley, J. F. *Information Circ. Econ. Geol. Res. Unit, Univ. Witwatersrand* **227**, (1990).
8. Atomic Energy Corporation Uranium in South Africa (Pretoria, 1985).
9. Phillips, G. N. & Myers, R. E. *Econ. Geol.* **6**, 598–608 (1989).
10. Allsopp, H. L. & Welke, H. J. in *Mineral Deposits of Southern Africa I* (eds Anhaeusser, C. R. & Maske, S.) 495–496 (Geological Society of South Africa, Johannesburg, 1986).
11. Krogh, T. E. *Geochim. cosmochim. Acta* **37**, 485–494 (1973).
12. Krogh, T. E. *Geochim. cosmochim. Acta* **46**, 637–649 (1982).
13. Corfu, F. *Contrib. Miner. Petrol.* **98**, 312–325 (1988).
14. Davis, D. W. *Can. J. Earth Sci.* **19**, 2141–2149 (1982).
15. Drennan, G. R., Robb, L. J., Meyer, F. M., Armstrong, R. A. & de Bruijn, H. S. *Afr. J. Geol.* **93**, 41–53 (1990).
16. Stacey, J. S. & Kramers, J. *Earth planet. Sci. Lett.* **26**, 207–221 (1975).
17. Key, R. M. & Wright, E. P. *J. geol. Soc. Lond.* **139**, 109–126 (1982).
18. Aldiss, D. N., Tombale, A. R., Mapeo, R. M. B. & Chiepe, M. *Bull. Geol. Surv. Dept Botswana* **33**, (1989).
19. Young, R. B. *The Blanket: A Study of the Auriferous Conglomerates of the Witwatersrand and the Associated Rocks* (Gurney & Jackson, London, 1917).

ACKNOWLEDGEMENTS. We thank J. Mortenson and J. Hodgson for reviews, and the Foundation for Research Development and the Anglo-American Corporation for financial and logistic support.

Causes of recent Himalayan uplift deduced from deposited patterns in the Ganges basin

Douglas W. Burbank

Department of Geological Sciences, University of Southern California, Los Angeles, California 90089-0740, USA

THE hypothesis¹ that late Cenozoic uplift of many of the world's mountain ranges was a response to climate change rather than to tectonic processes is based on the premise that enhanced erosion in an altered climate led to a reduction in both the total mass of a mountain range and its mean elevation. This reduction was coeval with increased regional relief and with passive isostatic uplift of the remaining summits. The hypothesis is difficult to evaluate because both climate and tectonic forces affect uplift in many active ranges. In the absence of sufficient denudational and uplift data from the mountains themselves, depositional patterns in adjacent basins may sometimes be used to distinguish between uplift arising from erosional or from tectonic processes. In foreland basins, uplift of the proximal basin and a reduction or absence of asymmetric subsidence are predicted responses to erosionally driven uplift². Here I show that patterns of sediment accumulation and the position of major rivers in the Neogene Gangetic foreland basin seem to have changed markedly in Plio-Pleistocene times. These changes lend support to the hypothesis that the importance of erosional unloading has increased in the Himalayas during the past 4 Myr.

The record of considerable climate change leading to widespread glaciation during late Cenozoic times is unambiguous in many mountain ranges³. It is likely that repeated alpine glaciation led to enhanced rates of alpine erosion, but there is an absence of reliable data with which to calibrate changes either in the mean altitudes of uplifting ranges or in past sediment fluxes out of them. Here I examine the depositional patterns in basins that bound uplifting mountains and within which sediment eroded from nearby ranges was trapped. Development of

two endmember models, one in which active tectonic shortening leads to mountain uplift and the other in which erosionally driven isostatic recovery leads to mountain uplift, can be used to predict contrasting depositional geometries. In the model, a mountain range adjacent to a foreland basin (Fig. 1) is underthrust by crust that behaves elastically, such that changes in the mass of the mountains cause flexural responses in the crust underlying the adjacent foreland basin^{2,4–6}. In the foreland basin itself, there are two types of drainage systems: transverse systems draining perpendicular to the mountains and longitudinal systems flowing parallel to them⁷. The model predicts that the relative importance of transverse relative to longitudinal drainages can be interpreted in terms of erosion-driven against thrust-driven uplift.

If thrusting is the primary cause of mountain uplift, the resultant crustal thickening causes asymmetric subsidence within the adjacent foreland (Fig. 1a), such that considerably more subsidence and deposition occurs in the proximal parts of the foreland^{4–6}, and this subsidence inhibits the progradation of transverse fans across the basin. Transverse rivers are therefore restricted to proximal areas, and longitudinal or axial rivers flow along the medial part of the foreland near the fan toes (Fig. 1c). In cross-section, the basin will appear to be underfilled⁵, because the rate of subsidence exceeds the rate of sediment supply to the transverse fans.

If enhanced erosion is the primary cause of uplift, isostatic recovery reduces the crustal root, and flexural uplift occurs across the foreland in response to the reduced mountain load (Fig. 1b). Enhanced erosion, however, leads to an increased flux of sediments into the basin, and this distributed sedimentary load, if preserved within the basin, both shifts the centre of loading away from the hinterland⁵ and increases the width of the foreland basin. Along with possible uplift of the proximal foreland basin, subsidence across the foreland is less asymmetric because of the distributed sedimentary load. Reduced subsidence in the proximal foreland and enhanced flux of sediments into the basin cause transverse fans to expand and longitudinal rivers to be displaced toward the distal basin (Fig. 1d). These models suggest that, despite a paucity of uplift and erosion data from a given mountain range, we can use the presence or absence

of asymmetric subsidence, a tabular rather than wedge-shaped geometry of sedimentary strata, and the relative importance of longitudinal or transverse rivers within an adjacent foreland basin, to differentiate between tectonically driven and erosionally driven uplift.

Depositional geometries within the Himalayan foreland basin can be evaluated in the context of these models to suggest which process is primarily responsible for late Cenozoic uplift of the Himalaya. The Indo-Gangetic foreland basin is developed in front of the transpressional ranges of western Pakistan and the collisional ranges of the Himalaya and Karakoram, and its strata onlap the western and northern margins of the Indian sub-continent. The huge volumes of sediment stored in the Bengal and Indus fans^{8,9} indicate that, throughout the past 20 Myr, the foreland has been nearly continuously filled to overflowing, such that all available sediment accommodation space was occupied. Assuming a relatively homogeneous rheology for the Precambrian crust of the Indian sub-continent¹⁰, the flexural wavelengths (~300 km) of the Indus and the Gangetic forelands should be similar. There is, however, a strong contrast in the modern drainage geometries observable in these two basins (Fig. 2a). Once it enters the foreland in northern Pakistan, the Indus River is rarely more than 100 km from the mountain front, and in many cases, such as with respect to the Sulaiman and Kirthar ranges, it flows longitudinally very close to the range front despite the presence of a broad, largely undeformed foreland in the medial and distal portions of the basin. These geometries seem to result from rapid asymmetrical subsidence and/or a relatively low sediment flux from the transverse fans. Except where the Indus drainage flows within the Hindu Kush and Himalaya, most of its bounding mountain ranges are relatively low in elevation (<2 km), are located within a semi-arid to arid climate regime, and have experienced little or no glacial erosion. These factors¹¹ should combine to provide a smaller transverse sediment flux into the middle and lower Indus River than in its course within the Himalaya.

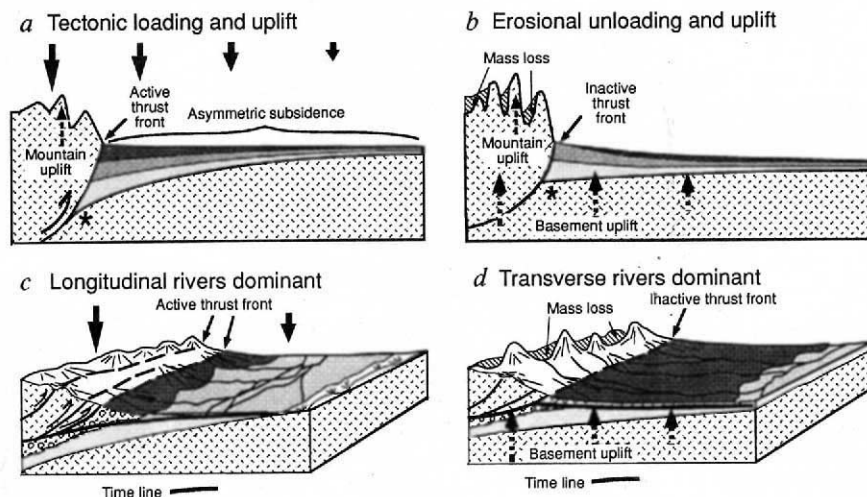
A strongly contrasting scenario is seen in the Ganges foreland basin (Fig. 2a). Except in the headward part of the drainages, the Yamuna and Ganges rivers form a longitudinal system flowing parallel to, but 200–300 km in front of, the mountain front. For much of their courses, these rivers are in the most distal part of the foreland basin, often along the feather edge. Although the foreland strata thicken to more than 6 km adjacent

to the mountain front, the foreland basin strata beneath the modern Ganges and Yamuna rivers are typically <1 km in thickness. Large, low-gradient transverse fans dominate the proximal and medial parts of the Ganges foreland basin. The progradation of these transverse fans has pushed the Ganges River into its present, very distal position. The cross-sectional asymmetry of the Neogene Ganges foreland fill¹² indicates that thrust loading has been the primary long-term (pre-Pliocene) control on basin subsidence (Fig. 2). The present drainage pattern and the cross-sectional geometry of the Plio-Holocene basin fill (Upper Siwalik: Fig. 2b and c), however, support a model in which erosional unloading has propelled late Cenozoic uplift of both the range and the proximal foreland basin, and in which increased sediment fluxes into the foreland have led to extensive progradation of more tabular transverse fans, as opposed to the prismatic geometries that prevail during times of asymmetries subsidence (Lower and Middle Siwalik, Fig. 2b and c).

The anomalous nature of the modern situation can be discerned through comparison with the geometries that prevailed during middle and late Miocene times, when the ancestral Indus River in northeastern Pakistan flowed east-southeast into the Ganges drainage^{7,13} and was part of a longitudinal river system that flowed along virtually the entire front of the Himalayan range. Although erosion or overthrusting of Miocene rocks precludes description of the drainage patterns in the most proximal part of that basin, preserved fluvial strata clearly indicate that the ancestral Indus–Ganges system flowed in the medial part of the basin and was often located far (>100 km) from the distal edge of the basin^{7,14}. Unlike the modern Ganges River, this ancestral river typically flowed along the surface of >2 km of previously accumulated strata in a foreland basin characterized by subsidence that was strongly asymmetric toward the hinterland. Synchronous intervals of accelerated subsidence^{15–17} suggest that thrust loading in the hinterland, probably including major basement-involved thrusts, were the primary controls on the subsidence and the basin-filling history during Miocene times.

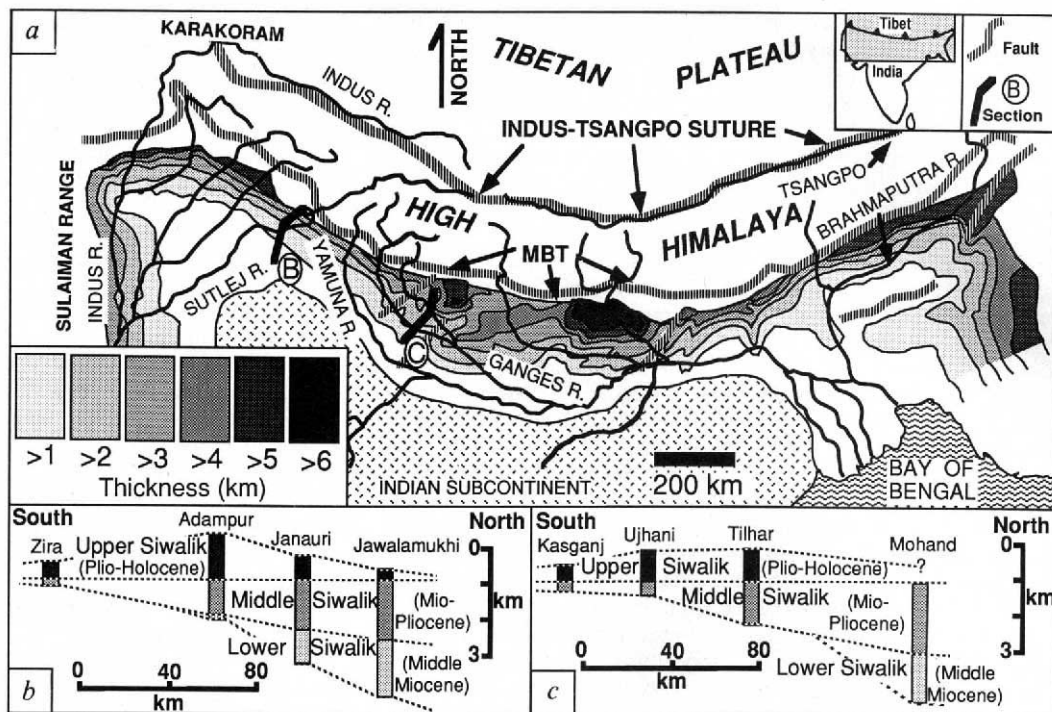
Additional information comes from tectonic, stratigraphic and geomorphological data. First, because the rate of Indo-Asian convergence has not slowed significantly since Miocene times, any diminution in the rate of thrust loading within the Himalaya must be attributed to changes in the regional partitioning of

FIG. 1 Schematic contrasts in subsidence and depositional patterns in foreland basins adjacent to mountains undergoing either tectonic loading or erosional unloading. Both scenarios cause uplift of individual summits, but during erosional unloading the mean surface elevation decreases. *a*, Tectonic loading due to crustal compression and thickening drives both uplift of the mountain summits (dashed arrow) and asymmetric subsidence (heavy arrows) in the foreland basin above an elastically loaded crust, such that wedge-shaped depositional units thicken toward the thrust front (* marks reference point at top of basement). *b*, Erosional unloading causes isostatic uplift of remaining peaks as the total mass of the mountains is reduced. Topography of the tectonically generated mountains (from *a*) is superimposed (hachured area) to show the net loss of mass despite summit uplift. Because of the elastic crustal response, the proximal foreland basin is also uplifted (heavy dashed arrows indicate basement uplift) and may be eroded. In cross-section, new sedimentary deposits are tabular, rather than wedge-shaped. *c*, Depositional patterns during active thrust loading are characterized by short transverse rivers and dominant longitudinal rivers flowing in the distal and medial parts of the foreland, often above several kilometres of previously deposited fluvial strata. Stratal time line illustrates asymmetric subsidence. *d*, During



the erosional phase, there is a net mass loss from the mountains (hachured area represents eroded topography of Fig. 1c), widespread basement uplift occurs, transverse rivers prograde across the foreland, and the longitudinal river is restricted to a distal position above thin to negligible previously deposited foreland strata.

FIG. 2 a, Major drainage patterns in the Himalayan system and isopach map of foreland strata (in kilometres of preserved thickness). An asymmetric wedge of sedimentary rocks thickens toward the mountain front along the entire range. After entering the foreland south of the main boundary thrust (MBT), the Indus flows as a longitudinal river parallel and close to the front of the Sulaiman and Kirthar Ranges. In contrast, after flowing as transverse rivers across the thick, pre-Pleistocene deposits of the Ganges foreland, the Yamuna and Ganges become longitudinal rivers restricted to the extreme distal margin of the depositional basin, a position consistent with transverse-fan progradation in response to erosional unloading in the main Himalayan Range. Major faults are shown by vertically hachured lines. Locations of cross-sections (b, c) are shown by heavy solid lines. b, c, Schematic transverse cross-sections of the Ganges foreland based on drill-hole data²⁹ showing stratal thickening toward the mountains during Middle and Lower Siwalik deposition, and stratal



thickening away from the mountains and more tabular depositional geometries during Upper Siwalik deposition from the Pliocene until present. Location of sections is shown by labelled lines in a.

shortening. Accelerated Plio-Pleistocene deformation in the Tien Shan and northeastern Tibet¹⁸⁻²¹ may be accommodating some of the convergence that was previously absorbed by the Himalaya. Second, although there is clear evidence of Pleio-Pleistocene shortening in the Himalaya^{12,22,23}, rates of denudation nearly equivalent to the rate of hanging-wall uplift²⁴ have often reduced the expected thrust loads within rapidly deforming areas. Third, modern levelling studies in the Nepal Himalaya²⁵ show that, despite ongoing uplift in the High Himalaya and the proximal foreland, the mean rate of erosional lowering at present exceeds the rate of interseismic uplift. Finally, during the Pliocene epoch at a time of accelerating uplift in the outer ranges of the Himalaya²² and rapidly changing climate²⁶, the longitudinal drainage system (the ancestral Indus River) in the north-western Himalaya was displaced by a transverse system (the ancestral Jhelum River)^{13,27}. This change was partially in response to relative uplift of the proximal foreland and an outward displacement of the depocenter^{16,28} and was accompanied by a grain-size coarsening and possibly an enhanced sediment flux out of the Himalaya.

The timing of the transition to a Gangetic foreland dominated by transverse rivers is poorly constrained. In northern Pakistan, it is younger than ~4 Myr (refs 13, 22, 27, 28), but the transition can only be defined as occurring during Upper Siwalik (Plio-Pleistocene) deposition in India^{14,29}. Nevertheless, these ages post-date by ≥ 4 Myr the estimated time when the Tibetan Plateau achieved nearly its present height ~8 Myr (ref. 30) and the apparently related initiation of the Indian monsoon³¹. Although considerable time-lag could be invoked to correlate the river-pattern transition with these tectonically driven events, it seems more closely linked in time to the initiation of widespread Northern Hemispheric continental and alpine glaciation during late Pliocene times^{3,26,32}.

This analysis shows that simple models predict very different responses for the depositional systems in foreland basins adjacent to ranges in which summits are uplifting because of thrust thickening and ones in which uplift is in response to erosion and isostatic recovery. Although the relative importance of

erosionally driven and tectonically driven uplift today or in the past is not precisely known, the data from the modern Indus and the Miocene Indo-Gangetic forelands suggest that thrust loading in the hinterland generated asymmetric subsidence, inhibited progradation of transverse fans and permitted longitudinal river systems to flow in medial or even proximal parts of the foreland basin. Despite evidence of active thrusting and local crustal thickening in parts of the Himalaya today, the modern Gangetic foreland is dominated by transverse drainages. Large Plio-Pleistocene sediment fluxes combined with both broadly distributed, less asymmetric subsidence and uplift of the proximal foreland have led to extensive progradation of transverse drainage systems that displaced the Ganges River to the feather edge of the present foreland basin: a position that is consistent with erosion-driven uplift in the adjacent Himalaya, but would be highly anomalous if thrust-driven crustal thickening were the primary cause of uplift in those same mountains. Improved calibration of changes in Neogene to Quaternary sediment fluxes into the Indo-Gangetic foreland and in surface elevations within the Himalaya are needed to evaluate this hypothesis more completely. □

Received 15 January; accepted 8 May 1992.

1. Molnar, P. & England, P. *Nature* **346**, 29-34 (1990).
2. Heller, P. L., Angevine, C. L., Winslow, N. S. & Paola, C. *Geology* **16**, 501-504 (1988).
3. Sibrava, V., Bowen, D. Q. & Richmond, G. M. *Quaternary Glaciations in the Northern Hemisphere* (Quat. Sci. Rev. 5) (1986).
4. Flemings, P. B. & Jordan, T. E. *J. geophys. Res.* **94**, 3851-3866 (1989).
5. Flemings, P. B. & Jordan, T. E. *Geology* **18**, 430-434 (1990).
6. Angevine, C. L., Heller, P. L. & Paola, C. *Quantitative Sedimentary Basin Modelling* (American Association of Petroleum Geologists, Tulsa, 1990).
7. Burbank, D. W. & Beck, R. A. *Geol. Rund.* **80** (1991).
8. Naini, B. P. thesis, Columbia Univ. (1980).
9. Curry, J. R. *Geology* **19**, 1097-1100 (1991).
10. Molnar, P. A. *Rev. Earth planet. Sci.* **12**, 489-518 (1984).
11. Ahnert, F. *Am. J. Sci.* **268**, (1970).
12. Raiverman, V., Kunte, S. V. & Mukherjee, A. *Petrol. Asia J.* 67-92 (1983).
13. Reynolds, R. G. H. thesis, Dartmouth College (1980).
14. Tandon, S. K. & Rangaraj, S. in *Structural Geology of the Himalaya* (ed. Saklani, P. S.) (Today and Tomorrow's Printers and Publishers, Delhi, 1979).
15. Burbank, D. W. & Beck, R. A. *Geol. Bull. Univ. Peshawar* **22**, 11-24 (1989).
16. Burbank, D. W., Beck, R. A. & Talling, P. J. *Geol. Soc. Am. Abst. Prog.* **23**, 241 (1991).
17. Johnson, N. M., Stix, J., Tauxe, L., Cerveny, P. F. & Tahirkehi, R. A. *J. Geol.* **93**, 27-40 (1985).

18. Molnar, P., Burchfiel, B. C., K'uangyi, L. & Zhao, Z. *Geology* **15**, 249–253 (1987).
19. Molnar, P. *et al. Geol. Soc. Am. Abstr. Prog.* **23**, 373 (1991).
20. Burchfiel, B. C. *et al. Tectonics* **10**, 1091–1110 (1991).
21. Zhang, P. *et al. Tectonics* **10**, 1111–1129 (1991).
22. Burbank, D. W. & Reynolds, R. G. H. in *New Perspectives in Basin Analysis* (ed. Kleinspehn, K. L. & Paola, C.) (Springer, New York, 1988).
23. Karunakaran, C. & Rao, A. R. *Geol. Sur. India Misc. Publ.* **41**, 1–66 (1979).
24. Burbank, D. W. & Beck, R. A. *Geology* **19**, 1169–1172 (1991).
25. Adhikari, K. *et al. Eos Trans.* **72**, 496 (1991).
26. Liu, T. *Aspects of Loess Research* (China Ocean, Beijing, 1987).
27. Burbank, D. W., Beck, R. A., Reynolds, R. G. H., Hobbs, R., Takirkheli, R. A. K. *Geology* **16**, 1143–1146 (1988).
28. Reynolds, R. G. H. & Johnson, G. D. in *The Chronology of the Geologic Record* (ed. Snelling, N. J.) (Geological Society of London, 1985).
29. Acharyya, S. K. & Ray, K. K. *AAPG Bulletin* **66**, 57–70 (1982).
30. Harrison, T. M., Copeland, P., Kidd, W. S. F. & Yin, A. *Science* **255**, 1663–1670 (1992).
31. Prell, W. L. & Kutzbach, J. E. *Eos* **72**, 257–258 (1991).
32. Shackleton, N. J. *et al. Nature* **307**, 620–623 (1984).

ACKNOWLEDGEMENTS. I thank R. Beck and T. Muller for assistance in the field, and members of the University of Peshawar, particularly R. A. K. Tahirkheli and J. Khan, for support. The sedimentary isopach base map was provided by Bob Reynolds. This research was partially funded by the NSF and National Geographic Research.

A vascular conducting strand in the early land plant *Cooksonia*

D. Edwards, K. L. Davies & L. Axe

Department of Geology, University of Wales College of Cardiff,
PO Box 914, Cardiff CF1 3YE, UK

THE late Silurian–early Devonian genus *Cooksonia*, characterized by smooth isotomously branching axes and solitary, terminal sporangia¹, has long been regarded as the archetypal vascular plant because of its age and simplicity of organization. The discovery of stomata, sterome² and thick-walled spores^{1,3} in *Cooksonia pertoni* and *C. hemisphaerica* confirmed its land-plant status, but tracheids have never been demonstrated in attached axes⁴. Here we report on tubes with differentially thickened walls typical of tracheary elements found in the central region of axes of Lower Devonian unequivocal *C. pertoni*, vindicating Lang's belief that *Cooksonia* was a vascular plant¹.

The specimens come from the same stream section through Lower Old Red Sandstone fluviatile sediments on Brown Cleve Hill, Shropshire, that have yielded the well preserved *C. pertoni*^{2,3}. Spore assemblages indicate a Gedinian age (middle *micromnatus-newportensis* miopore assemblage subzone)⁵. The fossils are coalified, partially compressed, showing remarkable preservation of axial and sporangial cells, and occasionally *in situ* spores. They were recovered from the residue left after dissolving the grey silty matrix with hydrofluoric acid. After secure identification involving scanning electron microscopy, two specimens were prepared for sectioning. Axes were separated from sporangia, treated with fuming nitric acid for 30 min, washed thoroughly and dehydrated in 100% acetone. After passing through increasing ratios (1:3, 1:1, 3:1) of low-viscosity Spurr resin to acetone (modified from ref. 6), they were embedded in Spurr resin⁷ and polymerized for 24 h at 70 °C. Specimens were sectioned using an LKB Ultratome 8801A with glass knives and 60–90 nm-thick sections (UF597 series) were collected on collodion-coated copper grids and stained with potassium permanganate, uranyl acetate and lead citrate (modified from ref. 6) before transmission electron microscopy using a JEOL 100S at an accelerating voltage of 80 kV. For light microscopy, sections of 1 µm thickness (UF1877 series) were stained with 0.25% (w/v) toluidine blue in 0.25% (w/v) sodium tetraborate solution, air-dried, and mounted in Elvacite.

Figure 1a shows a stalked sporangium with a shape typical of *C. pertoni*³. The well-preserved *in situ* spores belong to the dispersed taxon *Streelispore newportensis* (Fig. 1d) and identify the specimen as *C. pertoni* ssp. *apiculispore*³. The sporangium has a well defined rim with numerous papillate cells (Fig. 1c), not previously recorded. The subtending axis, although short

and apparently shrivelled, shows at least eight stomata (Fig. 1b), but remaining epidermal cells are indistinct. Its fractured end reveals fragmentary cell walls and abundant isolated pyrite octahedra that disrupt and distort cellular organization. Towards the centre of the axis are fragments of 2–3 cells with thickenings similar to those in the xylem of tracheophytes (Fig. 2b). Where intact, the external wall may be smooth or corrugated, depressions marking the positions between secondary thickenings. The latter appear annular with some interconnections (Fig. 2b). A fractured portion of the same tracheid shows a cast of the lumen, indentations marking the positions of the secondary wall, and also the inner surface of the tracheid wall with thickenings (Fig. 2a). At high magnifications, the coalified walls appear granular, but lack indications of further pitting as, for example, characterizes *Gosslingia*⁸. Owing to their incomplete nature, it cannot be decided whether these elements with differentially thickened walls are tubes or tracheids, but the latter seem most likely. Dimensions are given in Table 1. Longitudinal sections through fragments of tracheids have been observed by transmission electron microscopy (Fig. 3a–b) and by light microscopy (Fig. 3c–f). The best example of the former (Fig. 3b)

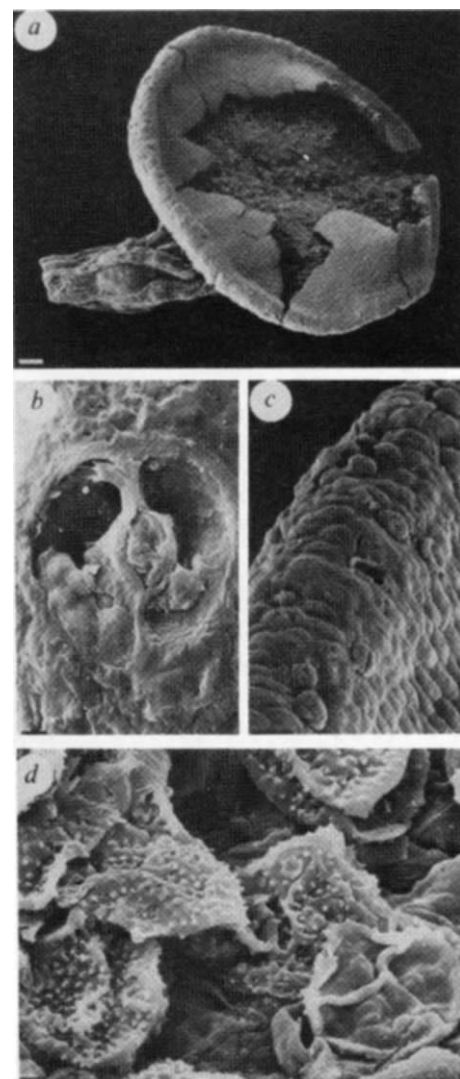


FIG. 1 Scanning electron microscopy (SEM; Cambridge microscope S360) of *Cooksonia pertoni* ssp. *apiculispore*, Gedinian Shropshire (DE41). a, Complete specimen with short stalk (scale bar, 100 µm); b, stoma (scale bar, 5 µm); c, sporangial rim with papillate cells (scale bar, 30 µm); d, *in situ* spores from the dispersed taxon *Streelispore newportensis*, although this specimen lacks the characteristic proximal folding (scale bar, 5 µm).

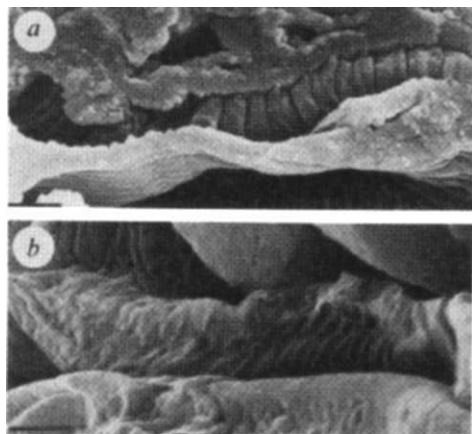


FIG. 2 SEM of tracheids (DE41); scale bars, 1.0 μm . *a*, Cast of lumen with indentations marking positions of secondary thickenings: thickenings themselves are just visible on extreme left; *b*, outer wall of tracheid with ridges marking positions of thickenings.

shows a longitudinally but obliquely cut tube with diameter, including walls, of 3.2 μm . The secondary thickenings appear as circular to wedge-shaped outgrowths of the longitudinal walls, with width of attachment less than their maximum thickness. The distance between adjacent, opposite thickenings (pitch) is 2.16–2.66 μm , but whether or not the thickenings were tilted annular or spiral cannot be distinguished. The band (X in Fig. 3*a*) may represent a thickening cut parallel to the primary wall. Difficulties were encountered in observing, interpreting and photographing sections in light microscopy owing to the small size of cells and thickenings. Those in Fig. 3*e* hint at having annular thickenings and are less compressed than the surrounding presumed cortical tissues (Fig. 3*c*).

Thus three specimens of *C. pertoni* have yielded tracheids: one identifiable to subspecies level, the two remaining lacking the critical evidence from *in situ* spores. Variation in the proximity of adjacent thickenings may be infraspecific, but most likely relates to distance below the sporangium and the extent of axis elongation following differentiation. The scanning electron

TABLE 1 Quantitative data for Geddinnian *C. pertoni*

Character	Specimen reference no.		
	UF1877	UF597	DE41
Length of axis (excluding sporangium) (μm)	705	1,400	666
Width of axis immediately below sporangium (μm)	314	514	414
Width of proximal end of axis (μm)	112	171	190
Width of sporangium (μm)	630	857	1,560
Height of sporangium (μm)	171	200	269
Width of axis at location of conducting tissue (μm)	112–116	—	215
Number of conducting cells visible across diameter of axis	2–3	—	—
Length of conducting tissue visible in section (μm)	208–232	21	—
Width of xylem (μm)	14–16	—	—
Area of axis occupied by xylem (%)	1.45–2.04	—	—
Width of conducting cells (μm)	5–11	3.2–7.0	2.0
Width of primary cell wall (μm)	0.5–2.0	0.26–0.53	0.37
Width of secondary cell wall (μm)	—	0.53–1.40	—
Maximum height of secondary thickening (μm)	—	0.76–2.28	—
Distance between adjacent thickenings (μm)	—	2.0–3.6	0.39

micrographs are closer to the sporangium, whereas observations by light and transmission electron microscopy were more distal, possibly below the level of extension growth. The number of tracheids in stelar cross-section is similar and small, but the nature of the preparations made them difficult to count. Their small number may be due to poor preservation, but probably reflects the original state, as is seen in small axes of *Rhynia gwynne-vaughanii*. This would account for our difficulty in finding tracheids in the several hundred specimens examined solely by scanning electron microscopy.

Such evidence demonstrates the tracheophyte affinity of *C. pertoni* in the early Devonian, but plants of similar morphology also occur in the Silurian, the oldest being late Wenlock (*lundgreni* graptolite Biozone^{4,9})—the earliest evidence for fertile pteridophyte-like plants. The question thus arises as to the vascular status of these older representatives. Central strands

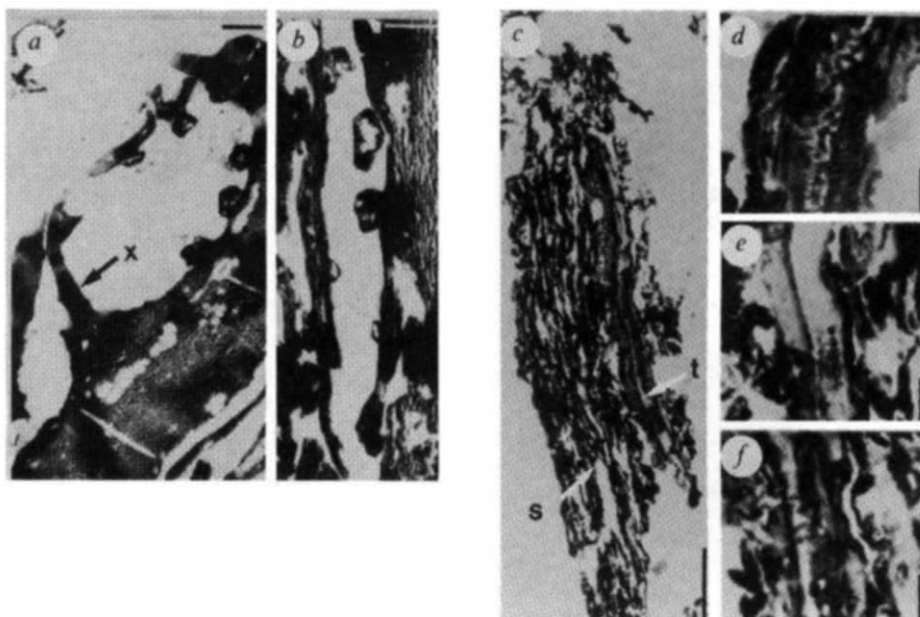


FIG. 3 *a* and *b*, Transmission electron microscopy of tracheids (UF597 series), with secondary thickenings in transverse section. 'X' may be a thickening cut parallel to the tracheid wall. Scale bars, 2 μm . *c*–*f*, Light microscopy of axis in longitudinal section of *C. pertoni* (UF1877 series). *c*, Part of axis with sterome 's' and tracheids 't' (indicated by arrows); scale bar, 50 μm . *d*–*f*, Fragments of presumed tracheids with secondary thickenings; scale bar, 5 μm .

of similar dimensions in Pridoli *C. pertonii*^{4,10} may be interpreted as tracheidal because, on anatomical grounds, they are closely related to Gedinnian forms³. Extension of this inference to Ludlow and Wenlock cooksonias without anatomical preservation is less secure. Perhaps the earliest representatives lacked vascular tissues and these were acquired sequentially, as the final phase of homoiohydry¹¹. It is also possible that Silurian central strands were of similar construction to those in *Aglaophyton major*¹². Indeed a recent report tentatively united *Cooksonia* (then presumed to lack tracheids) with *Aglaophyton* in a pro-tracheophyte grade¹³, but whether plants with moss-like conducting tissues are a sister group or are ancestral to vascular plants must remain unanswered because of insufficient evidence. □

Received 30 March; accepted 1 May 1992.

- Lang, W. H. *Phil. Trans. R. Soc.* **B227**, 245–291 (1937).
- Edwards, D., Fanning, U. & Richardson, J. B. *Nature* **323**, 438–440 (1986).
- Fanning, U., Richardson, J. B. & Edwards, D. *Evol. Trends Pl.* **2**, 13–24 (1988).
- Edwards, D., Bassett, M. G. & Rogerson, E. C. W. *Lethaia* **12**, 313–324 (1979).
- Richardson, J. B., Ford, J. H. & Parker, F. J. *Micropalaent.* **3**, 109–124 (1984).
- Taylor, W. A. & Taylor, T. N. *Am. J. Bot.* **74**, 437–443 (1987).
- Spurr, A. R. *J. Ultrastruct. Res.* **26**, 31–43 (1969).
- Kenrick, P. & Edwards, D. *Bot. J. Linn. Soc.* **97**, 95–123 (1988).
- Edwards, D., Feehan, J. & Smith, D. G. *Bot. J. Linn. Soc.* **86**, 19–36 (1983).
- Richardson, J. B. & Edwards, D. in *A Global Standard for the Silurian System* (eds Holland, C. H. & Bassett, M. G.) 216–226 (National Museum of Wales, Cardiff, 1989).
- Raven, J. A. in *Advances in Botanical Research* (ed. Woodhouse, H. W.) 153–219 (1977).
- Edwards, D. S. *Bot. J. Linn. Soc.* **93**, 173–204 (1986).
- Kenrick, P. & Crane, P. R. *Bot. Gaz.* **152**, 335–356 (1991).

ACKNOWLEDGEMENTS. Research supported by grants from the NERC. We thank U. Fanning for assistance in earlier stages of the project.

Do simple rules apply in honey-bee nestmate discrimination?

Michael D. Breed & Glennis E. Julian

Department of Environmental, Population and Organismic Biology, and Institute for Behavioral Genetics, The University of Colorado, Boulder, Colorado 90309-0334, USA

How do honey-bees integrate cues in nestmate recognition? Honey-bees are able to discriminate nestmates from non-nestmates using olfactory cues or contact chemoreception. Nestmates may differ from non-nestmates in either cues produced by the individuals, which have a genetic basis, or in cues that are acquired from the hive environment. Recognitive cues are learned by worker bees early in adult life; the learned cue profile serves as a template for comparison and determination of the colonial membership of bees that are encountered. We report here the response of bees to two cues: hexadecane and methyl docosanoate.

There are several models of template formation that explain how social insects might integrate a number of cues to discriminate kin from non-kin (or nestmates from non-nestmates)^{1,2}. Getz¹ used simple allelic modelling to show that social insects could base discrimination either on the presence of shared cues (habituated-label acceptance) or on the presence of unexpected cues (foreign-label rejection). Breed and Bennett² formulated a decision rule based on an overriding cue used by colony members to discriminate nestmates from non-nestmates (the common feature model); this model is appropriate for cues that do not represent the alleles carried by each individual.

Breed and Stiller³ identified two chemical cues used in honey-bee group membership discrimination. When bees are treated with these cues, hexadecane and the methyl docosanoate, they are perceived as having different discrimination cues than those of untreated sister bees. Hexadecane and methyl docosanoate

TABLE 1 Results of transfer of individuals from donor to recipient groups

	Number of attacks	N	Attacks (%)	
Controls	81	232	34.9	a
Single cue into unlike cue	60	98	61.2	b
Untreated with single cue	110	204	53.9	b
Hexadecane into combination	17	50	34.0	a
Methyl docosanoate into combination	29	50	58.0	b
Combination into hexadecane	20	56	36.0	a
Combination into methyl docosanoate	38	57	67.0	b

N is the number of individuals transferred and the number of attacks is the number of individuals bitten or stung after introduction. The letters at the end of each row show the presence or absence of statistically significant differences between tests; all types of tests followed by the same letter do not differ significantly and tests followed by different letters do differ significantly ($P < 0.05$). The treatments were: controls, donor bees and recipient bees that had been exposed to the same treatment. The four control groups were, untreated bees into untreated, hexadecane-treated bees into hexadecane recipients, methyl docosanoate-treated bees into methyl docosanoate recipients, and also bees treated with both methyl docosanoate and hexadecane into hexadecane/methyl docosanoate-treated bees. Single cue into unlike cue, bees treated with hexadecane into methyl docosanoate-treated bees, and vice versa. Untreated with single cue, hexadecane-treated bees into untreated bees; untreated bees into hexadecane-treated bees; methyl docosanoate-treated bees into untreated; untreated bees into methyl docosanoate-treated bees. Single cue into combination, hexadecane-treated bees introduced into bees treated with both hexadecane and methyl docosanoate; and methyl docosanoate-treated bees introduced into the combination of both compounds. Combination into single cues, bees treated with both compounds introduced into bees with either hexadecane or methyl docosanoate.

are similar to compounds found in beeswax and in honey-bee cuticular hydrocarbons, but are absent or are present in only trace amounts in honey-bee colonies⁴. Although hexadecane and methyl docosanoate both affect nestmate discrimination, the previous work³ did not address whether or how bees put the cues in order of priority when both are present.

Breed^{5,6} established the experimental methodology used in these experiments. Bees were removed from a wax comb of pupae within 24 hours of adult emergence and placed in cardboard containers. In the experiments reported here each container held 10 bees. Of a set of 11 such containers, one is designated the donor container and 10 are designated recipients. When the bees were 5 days old, each bee was removed from the donor container and placed into one of the recipient containers. The behavioural interactions between the donor bee and recipient bees were observed for 5 min. The observer was 'blind' with respect to the treatment that bees in the replicate might have received. If the donor bee is bitten or stung, the replicate is scored as a rejection (aggression having occurred) whereas if the donor bee is not bitten or stung the replicate is scored as an acceptance (no aggression). No donor bee was used more than once and no recipient container was used more than once; this procedure controls for transfer of discrimination labels among bees and for learning of labels by bees in the recipient containers during trials.

The bees in the donor and recipient groups were from the same comb and consequently were at least half-sisters (shared the same mother) and possibly supersisters (shared both mother and father). Bees in some groups received chemical treatment; either 0.1 g methyl docosanoate or 100 µl hexadecane were placed on a 5.5 cm filter paper disk. This disk was then placed in the container of bees for the 5 days before testing. These quantities were known from previous experiments^{3,7} to have similar effects on discrimination. In some of the tests the bees received both compounds; they were placed on a single paper disk. We refer to the groups of bees with both cues as 'combination' cues.

TABLE 2 Statistical comparisons among the behavioural tests in Table 1

	Grouped controls	Single/unlike	Untreated/single	Hexadecane/comb.	Docosanoate/comb.	Comb./hexadecane	Comb./docosanoate
Grouped controls	NA	19.49***	15.93***	NS	9.22**	NS	19.05***
Single/unlike		NA	NS	9.83**	NS	9.29**	NS
Untreated/single			NA	6.38*	NS	5.83*	NS
Hexadecane/comb.				NA	5.79*	NS	11.38***
Docosanoate/comb.					NA	5.28*	NS
Comb./hexadecane						NA	10.83***
Comb./docosanoate							NA

Column and row labels are abbreviations of the row labels in Table 1. Values are from 2×2 χ^2 contingency tables, 1 d.f.; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Bees were tested as described above, using all 14 possible variations of donors and recipients that can be generated by using no cue, one cue, or the two cues in combination. The results are shown in Table 1 and the statistical analysis of the data is in Table 2. Some of the combinations are grouped together in Table 1; in those cases there is no significant heterogeneity among the grouped treatments (for controls, $\chi^2 = 6.83$, d.f. = 3; for single cue into unlike cue, $\chi^2 = 1.17$, d.f. = 1; for untreated with single cue, $\chi^2 = 3.53$, d.f. = 3). This finding is particularly important, as it indicates that hexadecane and methyl docosanoate, separately, have indistinguishable effects. One cue does not have stronger effects than the other when they are tested independently.

Three of the treatment types had relatively low rejection rates (controls, hexadecane donor and combination recipients, and combination donors and hexadecane recipients). The other treatments had rejection rates above 50% and in some cases above 60%. The three treatment types with low rejection rates are not significantly different from one another, but they are each significantly different from the four treatment types with high rejection rates (Table 2), which also do not differ significantly from one another.

The combination treatments are of special interest. When hexadecane was the matching compound between the two groups, there was a low rate of rejection. When methyl docosanoate is the matching compound, either on the donor bee or in the recipient group, there is a high rate of rejection. When methyl docosanoate is the matching compound and hexadecane is present in either the donor or recipient group, methyl docosanoate did not affect the outcome; bees were treated as if their cues did not match. Hexadecane has a higher priority than methyl docosanoate as a cue.

What general rules of cue use can be formulated as a basis for a general understanding of honey-bee nestmate discrimination? First, the common feature model of nestmate discrimination is strongly supported by several studies^{7,8}; a single cue, such as hexadecane, spread through a colony, can override the effects of other cue differences. In unmanipulated colonies, common feature cues are transferred from the comb wax to bees, and these cues override individually variable cues⁸. Second, the presence or absence of each cue, rather than the relative concentrations of cues, seems to have the greatest importance. Finally, cues are drawn, seemingly at random, from sets of similar compounds; the use of one compound as a cue does not predict the use of similar compounds³. The choice of cue compounds may be driven by the prioritization system^{8,9}, which could operate on arbitrary rules. Our data suggest a relatively simple system of ordering of priority of cues by the bees and the use of a limited number of common features among nestmates for nestmate recognition. □

- Francis, B. R., Blanton, W. E., Littlefield, J. L. & Nunemaker, R. A. *Ann. ent. Soc. Am.* **82**, 486–494 (1989).
- Breed, M. D. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2635–2637 (1981).
- Breed, M. D. *Anim. Behav.* **31**, 86–91 (1983).
- Breed, M. D., Stiller, T. M., Blum, M. S. & Page, R. E. Jr *J. chem. Ecol.* (in the press).
- Carlin, N. F., Hölldobler, B. *Behav. Ecol. Sociobiol.* **19**, 123–134 (1986).
- Crosland, M. J. W. *Anim. Behav.* **37**, 912–919 (1989).

ACKNOWLEDGEMENTS. We thank L. Snyder and T. Smith for their comments on this manuscript. This research was supported by a grant from the Animal Behavior Program of the National Science Foundation (to M.D.B. and Robert E. Page Jr. Principal Investigators) and a site grant in the Research Experiences for Undergraduates program of the National Science Foundation.

Developmental regulation of NMDA receptor-mediated synaptic currents at a central synapse

Shaul Hestrin

Department of Physiology, Box 0444, University of California School of Medicine, San Francisco, California 94143, USA

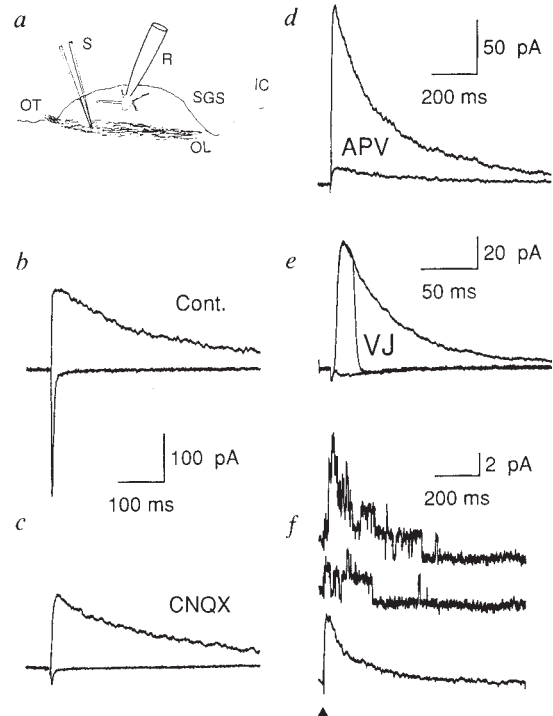
THE central nervous system has extraordinary plasticity in early life. This is thought to involve *N*-methyl-D-aspartate (NMDA) receptors¹ which, along with the non-NMDA receptors, mediate fast excitatory synaptic transmission². Although NMDA receptors may be transiently enhanced early in life^{3–6}, it has not been possible to demonstrate directly a functional change in the NMDA receptor-mediated synaptic response because of the voltage-dependence of the NMDA conductance and the overlapping inhibitory synaptic conductances. Here I report that the duration of evoked NMDA-receptor-mediated excitatory postsynaptic currents (e.p.s.cs) in the superior colliculus is several times longer at early developmental stages compared to that measured in older animals. In contrast, the amplitude of NMDA-receptor-mediated miniature e.p.s.cs does not change during development. The kinetic response of excised membrane patches to a brief activation of NMDA receptors is similar to that of the NMDA e.p.s.c, which suggests that the time course of the NMDA e.p.s.c. in the superior colliculus reflects slow NMDA channel properties as in the hippocampus^{7–9}. Therefore, these data indicate that the molecular properties of NMDA receptors are developmentally regulated and thus may be controlling the ability of synapses to change in early life.

Monosynaptic postsynaptic currents were evoked in neurons in superficial layers of the superior colliculus (stratum griseum superficiale) by applying brief electrical stimulation in the optic layer, which contains the afferent excitatory axons^{10,11} (Fig. 1a). The e.p.s.cs recorded at a holding potential of -80 mV exhibited a rapid onset and decay (Fig. 1b) and, at a more positive potential, a rapid onset but a slow decay (Fig. 1b). The non-NMDA antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX) specifically blocked the component observed at -80 mV (Fig. 1c). In contrast, the component observed at $+40$ mV was selectively blocked by the NMDA antagonist D-2-

Received 10 March; accepted 15 May 1992.

- Getz, W. M. *J. theor. Biol.* **99**, 585–597 (1982).
- Breed, M. D. & Bennett, B. in *Kin Recognition in Animals* (ed. Fletcher, D. J. C. & Michener, C. D., 243–285 (Wiley, Chichester, 1987).
- Breed, M. D. & Stiller, T. M. *Anim. Behav.* (in the press).

FIG. 1 *a*, Superior colliculus slices. R, Recording patch-pipette in the superficial layer stratum griseum superficiale (SGS); S, bipolar stimulating electrodes in the optic layer (OL); OT, optic tract; IC, inferior colliculus. *b*, Effect of membrane potential on the e.p.s.c. Current evoked by electrical stimulation recorded at -80 mV exhibits a fast rate of rise and a rapid decay. At a holding potential of $+40$ mV the decay of the e.p.s.c. was slow. 'Cont.' indicates control. *c*, Application of $10 \mu\text{M}$ CNQX diminished the response measured at -80 mV and reduced the rate of e.p.s.c. rise (10% to 90%) recorded at $+40$ mV. Application of CNQX changed the rate of rise from 4.0 ± 1.2 ms to 7.7 ± 2.2 ms ($n=6$), while having only small effect on the rate of decay. *d*, APV effects. The e.p.s.c. recorded from another cell in the presence of CNQX and in the presence of both CNQX and APV. *e*, Voltage-jump testing the adequacy of the voltage-clamp. The e.p.s.c. was measured at $+20$ mV and at 0 mV close to the null potential. Next, the responses to a voltage jump from $+20$ mV to 0 mV were recorded, with or without earlier stimulation. These two responses were subtracted and the resulting relaxation current was compared to the time course of the e.p.s.c. (marked VJ). Given that the relaxation current, which reflects the dynamic control of the postsynaptic membrane, was an order of magnitude faster than the e.p.s.c., it is reasonable to conclude that the e.p.s.c. is measured faithfully. *f*, Response of an excised patch to a brief glutamate application. Single responses to a 5 ms application of $200 \mu\text{M}$ glutamate ($\blacktriangle$) resulted in outward currents. The outside-out patch was excised from a 15-day-old preparation and held at $+40$ mV. Lower trace, average of 32 responses. METHODS. Both male and female Sprague Dawley rats were used. Parasagittal slices of the superior colliculus, $300 \mu\text{m}$ thick were obtained using vibratome¹². The slices were maintained at room temperature. The bath solution in all experiments, unless indicated otherwise, was (in mM): NaCl, 117; MgCl_2 , 4; CaCl_2 , 4; KCl, 4; NaH_2PO_4 , 1.2; NaHCO_3 , 26; Glucose, 15. The pipette solution was (in mM): Cs-gluconate, 122.5; CsCl, 17.5; HEPES (CsOH), 10; Na EGTA, 0.2; MgATP, 2; Na-GTP, 0.3; NaCl, 8. In some experiments the concentration of EGTA was raised to 10 mM. Whole-cell recordings were obtained from neurons in the superficial layer of the superior colliculus, (stratum griseum superficiale) using the 'blind' technique²³⁻²⁵. Stimulation was applied with bipolar stainless steel electrodes positioned at the optical layer. Fast inhibitory currents were blocked by GABA_A antagonist ($100 \mu\text{M}$ picrotoxin). Slow inhibitory synaptic currents, which increase potassium conductance, were blocked by replacing intracellular potassium ions with Cs. Whole-cell currents were recorded with a patch-clamp amplifier (EPC-7, List), filtered at 1 kHz and digitized at 2 – 10 kHz using a computer. Rapid application of glutamate to an outside-out patch of membrane was done



using a piezoelectric device²⁶ that displaced a flow pipe containing $200 \mu\text{M}$ glutamate. The glutamate containing solution as well as the bath solution contained CNQX ($2 \mu\text{M}$) and glycine ($10 \mu\text{M}$). The speed of the solution exchange was measured after the experiment by measuring the junction potential change at the tip of the pipette after removing the membrane patch. The rate of the solution exchange at the patch of membrane was rapid as indicated by the rate of Mg^{2+} block of NMDA channels activated in the absence of added Mg^{2+} (ref. 8). The onset and offset time of the solution exchange was less than 1.0 ms.

amino-5-phosphonovalerate (APV) (Fig. 1*d*). Thus the excitatory input to these cells is mediated by NMDA and non-NMDA receptors and is similar, pharmacologically and physiologically, to the dual component e.p.s.c. found in several other central nervous system (CNS) regions, including the CA1 region, of the hippocampus¹². To isolate NMDA-receptor-mediated e.p.s.c.s from other synaptic conductances, CNQX ($10 \mu\text{M}$) and picrotoxin ($100 \mu\text{M}$) were added routinely to the recording solution. The adequacy of the voltage-clamp measurements was tested using the voltage-jump technique (Fig. 1*e*).

To investigate whether the slow time course of the NMDA

e.p.s.c. could be either due to slow removal of transmitter or to slow intrinsic channel kinetics⁷⁻⁹, patches of membrane were excised from superior colliculus neurons in the outside-out configuration. Glutamate ($200 \mu\text{M}$) was briefly (4 – 8 ms; Fig. 1*f*) applied to the patch held at $+40$ mV in the presence of CNQX ($2 \mu\text{M}$). This resulted in outward single-channel currents (Fig. 1*f*). The ensemble current obtained by averaging 32 responses to a brief glutamate application resembled the e.p.s.c. (Fig. 1*f*; similar observations were obtained in 17 experiments). The similarity of the slow NMDA e.p.s.c. and the response of an excised patch to a glutamate pulse indicates that diffusion or

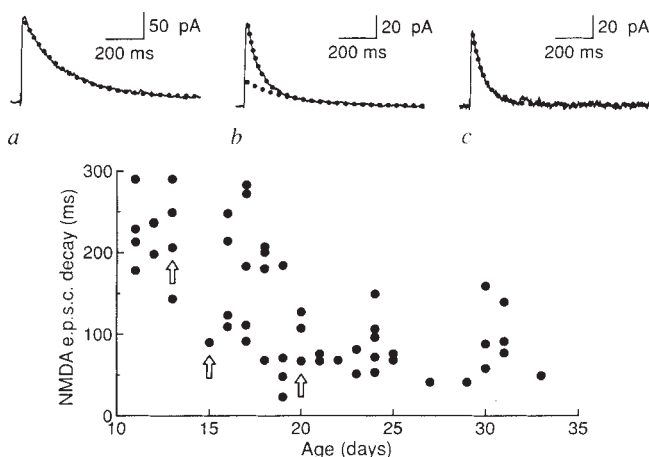


FIG. 2 Decay time of e.p.s.c.s recorded from 53 experiments. The membrane potential was held at $+40$ mV, in the presence of CNQX ($10 \mu\text{M}$) and picrotoxin ($100 \mu\text{M}$). Inset illustrates the e.p.s.c.s from three cells (ages 13, 15 and 20 days, indicated by the arrows). The dots in the inset represent the fitted single or double exponential functions. The time constants of the exponential functions were: *a*, 204.0 ms, 13-day-old animal; *b*, 226.4 ms (31%) and 54.9 ms (69%), 15-day-old; and *c*, 67.0 ms, 20-day-old. In most cells ($n=42$) the decay time (defined as time from the peak to 0.37 peak amplitude) was obtained by fitting a single exponential function to the average e.p.s.c. time course. When a single exponential function did not provide a good fit ($n=11$) the decay time of the NMDA receptor-mediated e.p.s.c. was measured directly as the time from the peak of the e.p.s.c. to 0.37 of the peak amplitude. For each cell at least five e.p.s.c.s recorded at $+40$ mV were averaged. Single or double exponential function was fitted to the average response using a least-sum-of-squares algorithm.

uptake do not determine the e.p.s.c. waveform and that single-channel properties alone can explain the NMDA e.p.s.c. time course⁸.

To compare the time course of the NMDA e.p.s.c. across different experiments, the e.p.s.c. decay time (defined as the time required for decay to 0.37 of peak amplitude) at a holding potential of +40 mV was measured in 53 experiments obtained at different post-natal ages (Fig. 2). Under these conditions, the decay time was 85 ± 37 ms (mean $\pm$ s.d., $n = 18$) for e.p.s.c.s of 23–33-day-old animals, significantly faster than the decay time of 213 ± 57 ms measured from 10–15-day-old animals. At an intermediate stage (16–22-day-old animals) the decay time was 127.2 ± 79 ms ($n = 23$). Although a single exponential function could fit most of the e.p.s.c.s (Fig. 2*a, c*), several e.p.s.c.s ($n = 11$) required two exponential functions to be fitted (Fig. 2*b*).

The current-voltage relationship of the NMDA-receptor-mediated e.p.s.c. has a region of negative conductance, resulting from a voltage-dependent Mg^{2+} block^{13,14}. It has been suggested that in the hippocampus^{15,16}, but not in the neocortex¹⁷, the Mg^{2+} blockade of the NMDA-activated conductance changes during development. In the superior colliculus the current-voltage relationship had a negative slope region both in young and in adult rats (data not shown), indicating that voltage-dependent blockade by Mg^{2+} is present at the developmental stages studied here. In the dentate gyrus the time course of the NMDA e.p.s.c. depends on membrane potential¹⁸. This was also seen in the superior colliculus: the decay time at -20 mV was

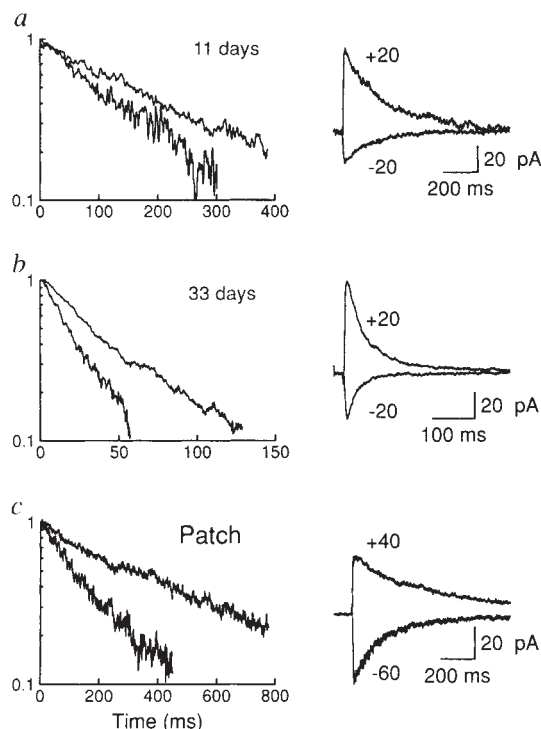


FIG. 3 Time course of the e.p.s.c. and of glutamate-induced NMDA-receptor-mediated responses is voltage dependent. *a, b*, Time course of the NMDA e.p.s.c. was recorded at +20 mV (outward current) and at -20 mV (inward current). Five responses were averaged at each synapse. Plot at the left is the normalized time course of the e.p.s.c.s recorded at a holding potential of +20 and -20 mV from an *a*, 11- and *b*, 33-day-old rat, plotted using semi-logarithmic axes. Plots at the right illustrate the time course of the e.p.s.c.s using linear axes. CNQX ($10 \mu M$) and picrotoxin ($100 \mu M$) were present. *c*, The response of an excised outside-out patch from a 15-day-old animal to a brief application of glutamate. The bath and glutamate containing solutions contained 0.5 mM Ca^{2+} and no added Mg^{2+} . Glutamate ($200 \mu M$) was briefly applied (5 ms; see Fig. 1 for details). The glutamate and bath solutions contained $10 \mu M$ glycine and $2 \mu M$ CNQX. The illustrated ensemble currents were averages of 16 responses at +40 mV (outward current) and -60 mV (inward current).

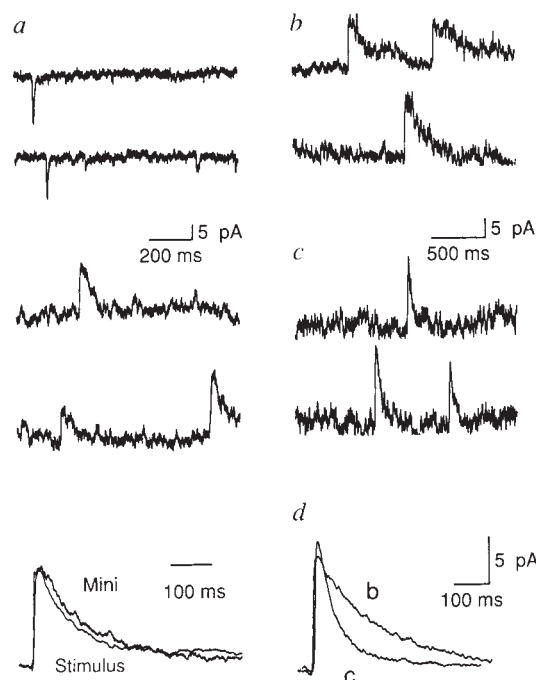


FIG. 4 Miniature NMDA-receptor-mediated e.p.s.c.s. *a* (top), Control solution without CNQX at a holding potential of -80 mV; middle, after the addition of CNQX at a holding potential of +40 mV; bottom, superposition of the normalized, average miniature e.p.s.c. and the evoked e.p.s.c. recorded at +40 mV in the presence of CNQX. *b*, Minis recorded in the presence of TTX ($1 \mu M$) from a 13-day-old rat. *c*, Minis recorded in the presence of TTX from a 17-day-old rat. *d*, Comparison of the average minis, same cells as in *b* ($n = 215$) and *c* ($n = 141$). The amplitudes and decay times were 12.2 pA and 186 ms (trace marked *b*); and 13.1 pA and 55.2 ms (trace marked *c*). Miniature e.p.s.c.s were recorded on video tape for off-line analysis. Minis were detected as a threshold crossing from a baseline. The threshold was set at +6 pA. The baseline was calculated for a 10 ms segment and the threshold crossing was measured at 10 ms ahead of the baseline segment. The baseline noise was 1.24 pA standard deviation for *b* and 1.19 pA for *c*. The average number of minis recorded was 136 ± 61 , $n = 5$.

about half of that measured at +20 mV. A change in the voltage sensitivity of the NMDA channel kinetics could underlie the developmental changes in the time course of the e.p.s.c. But the voltage sensitivity of the e.p.s.c. time course was observed at all ages (Fig. 3*a, b*): the average decay time was 211 ms at +20 mV compared with 120 ms at -20 mV for 10–13-day-old animals ($n = 5$), whereas the average decay time for 21–33-day-old animals was 77 ms at +20 mV and 33 ms at -20 mV ($n = 5$). Thus, the voltage-dependent properties of the NMDA-receptor-mediated e.p.s.c. seem to be present as early as 10 days after birth.

The voltage-dependency of the NMDA e.p.s.c. time course presumably reflects NMDA channel kinetics because transmitter removal is not likely to be affected by the membrane potential. Thus it would be expected that the response of excised patches to brief glutamate application would also show voltage-dependent kinetics. The time course of the response to a brief glutamate application was slow at +40 mV compared with that obtained at -60 mV (Fig. 3*c*, similar observations were made in four experiments). These experiments imply that changes in channel kinetics and not transmitter removal underlie the developmental change in e.p.s.c. time course.

The developmental change in the time course of the NMDA e.p.s.c. might reflect, in part, changes in the time course of transmitter release. To exclude this possibility, miniature spontaneous e.p.s.c.s (termed 'minis'), thought to represent release of transmitter packets at single synaptic contacts, were measured in the presence of tetrodotoxin (TTX) to block sodium-dependent action potential (Fig. 4*a* top, middle). The time course of an averaged NMDA mini exhibited a slow rise and a slow decay,

similar to the time course of the evoked e.p.s.c. measured before the addition of TTX (Fig. 4a bottom) indicating that the time course of the NMDA e.p.s.c. is not influenced by asynchronous release.

The average amplitude and decay time of NMDA minis were measured in animals at different ages (11–23 days old, $n = 5$). The range of average NMDA mini time courses was large (from 16 ms to 186 ms; $n = 5$; mean $\pm$ s.d., 95 ± 67 ms). In contrast, the range of the NMDA mini amplitudes was small (12.2 pA to 14.4 pA; $n = 5$; mean $\pm$ s.d., 13.5 ± 0.9 pA). Figure 4b, c illustrates minis obtained from 11- and 17-day-old rats. Further experiments are required to determine whether the NMDA minis have a binomial distribution representing the mature and the immature form. But the data obtained so far indicate that the amplitude of the NMDA minis are not correlated with the mini duration and therefore do not change during development.

The development of the NMDA e.p.s.c. reported here and the development of the end-plate current (e.p.c.) at the neuromuscular junction are very similar. At the neuromuscular junction the time course of the embryonic e.p.c. is slow compared with that of the adult¹⁹. The e.p.cs can be fitted with a single exponential function both at early and at late stages, but are biphasic at an intermediate stage²⁰, which is similar to the finding reported here (Fig. 2). At the neuromuscular junction, the biphasic time course of the e.p.c. is the result of mixed populations of channels mediating synaptic current, a similar explanation could underlie the biphasic NMDA e.p.s.c.. Because the time course of the NMDA-receptor-mediated e.p.s.c. is determined by the kinetic properties of NMDA-activated channels (refs 7–9; Figs 1 and 3) the changes in the time course of the NMDA e.p.s.c. that occur during development could reflect structural changes in NMDA-activated channels. By analogy with the NMJ development, the changes in the NMDA e.p.s.c. may reflect developmentally controlled genetic expression of the NMDA receptor subunits^{21,22}.

The important role of the NMDA-receptor-mediated e.p.s.c. in governing activity-dependent synaptic plasticity derives from its ability to allow calcium influx at synaptic sites. Thus, by changing the kinetics of the NMDA e.p.s.c. the level of activity required for inducing change would be affected. This model suggests that the immature NMDA receptor is 'permissive', requiring only moderate activity to induce the formation of specific neuronal wiring. Once these circuits have been laid down, the mature form of the NMDA receptor is expressed, increasing the activity threshold that has to be reached for further change. □

Rapid switching to multiple antigenic and adhesive phenotypes in malaria

David J. Roberts, Alister G. Craig,
Anthony R. Berendt, Robert Pinches, Gerard Nash*,
Kevin Marsh & Christopher I. Newbold

Molecular Parasitology Group, Institute of Molecular Medicine,
John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK

* Department of Haematology, University of Birmingham, Edgbaston,
Birmingham B15 2TT, UK

ADHESION of parasitized erythrocytes to post-capillary venular endothelium¹ or uninfected red cells^{2–4} is strongly implicated in the pathogenesis of severe *Plasmodium falciparum* malaria. Neo-antigens at the infected red-cell surface adhere to a variety of host receptors^{5–9}, demonstrate serological diversity in field isolates^{10,11} and may also be a target of the host-protective immune response¹². Here we use sequential cloning of *P. falciparum* by micromanipulation to investigate the ability of a parasite to switch antigenic and cytoadherence phenotypes. Our data show that antigens at the parasitized cell surface undergo clonal variation *in vitro* in the absence of immune pressure at the rate of 2% per generation with concomitant modulations of the adhesive phenotype. A clone has the potential to switch at high frequency to a variety of antigenic and adhesive phenotypes, including a new type of cytoadherence behaviour, 'auto-agglutination' of infected erythrocytes. This rapid appearance of antigenic and functional heterogeneity has important implications for pathogenesis and acquired immunity.

During studies on cytoadherence of infected erythrocytes we selected various phenotypes from a *P. falciparum* line IT 4/25/5. Parasites were selected for increased binding to C32 amelanotic melanoma cells, endothelial cells or rosetting of uninfected erythrocytes (Fig. 1). We noticed that selections for endothelial binding and rosetting were associated with changes in antigenic type. To investigate the relationship between antigenic and adhesive phenotypes and to study antigenic variation in the absence of selection pressure, we carried out a series of clonings by micromanipulation.

Four clones (A1–4) were derived from the endothelial binding line ITO4, three of which (A2–4) were antigenically indistinguishable from the parent. A4 was subcloned, yielding 21 'C' clones (Fig. 1). Cross-contamination was excluded by hybridization with the probe pC4.H32 (ref. 24). All clones gave an identical fingerprinting pattern which was different from that of other parasite lines (data not shown).

We analysed the antigenic phenotype of these clones (Fig. 2). Ten of the 21 C clones were of the same variant antigenic type as A4, but the remainder were antigenically distinct. Of six variant C clones analysed for antigenic similarity with each other, only two (C2 and C10) were antigenically indistinguishable, indicating that in *P. falciparum* antigenic variation is highly diverse (Fig. 2); it is also rapid. The proportion of variant C clones (11/21 or 0.52) reflects the antigenic heterogeneity within the A4 clone at the time of subcloning, provided cloning itself does not influence switching. Since 30 cycles elapsed from cloning A4 to the C subcloning, new antigenic types had appeared, on average, at 2.4% per generation, assuming equal growth rates for all variants. $((1 - r)^n = p$, where r is the switching rate per generation, n is the number of generations and p the proportion of the original antigenic type remaining after n generations.) This antigenic switching rate is of a similar order of magnitude to that recently observed in *Trypanosoma brucei*¹⁴.

Different isolates of *P. falciparum* adhere to CD36 and thrombospondin^{5–8}, intercellular adhesion molecule-1 (ICAM-1) (ref. 9) or rosette uninfected cells^{2,3}. Clone A4 binds to purified CD36 and ICAM-1 but does not rosette. Six C clones with the

Received 19 December 1991; accepted 11 May 1992.

- Constantine-Paton, M., Cline, H. T. & Debski, E. A. *Rev. Neurosci.* **13**, 129–154 (1990).
- Collingridge, G. L. & Lester, R. A. J. *Pharmacol. Rev.* **40**, 143–210 (1989).
- Tsumoto, T., Hagihara, K., Sato, H. & Hata, Y. *Nature* **327**, 513–514 (1987).
- Fox, K., Daw, N., Sato, H. & Czepita, D. *Nature* **350**, 342–344 (1991).
- Bode-Greuel, K. M. & Singer, W. *Dev. Brain Res.* **46**, 197–204 (1989).
- Kato, N., Artola, A., and Singer, W. *Dev. Brain Res.* **60**, 43–50 (1991).
- Hestrin, S., Sah, P. & Nicoll, R. A. *Neuron* **5**, 247–253 (1990).
- Lester, R. A., Clements, J. D., Westbrook, G. L. & Jahr, C. E. *Nature* **346**, 565–567 (1990).
- Gibb, A. J. & Colquhoun, D. *Proc. R. Soc. B* **243**, 39–45 (1991).
- Lund, R. D. & Lund, J. S. *Brain Res.* **42**, 1–20 (1972).
- Miyamoto, T., Sakurai, T. & Okada, Y. *Brain Res.* **518**, 166–172 (1990).
- Hestrin, S., Nicoll, R. A., Perkel, D. J. & Sah, P. *J. Physiol.* **422**, 203–225 (1990).
- Nowak, L., Bregestovski, P., Ascher, P., Herbet, A. & Prochianz, A. *Nature* **307**, 462–465 (1984).
- Mayer, M. L., Westbrook, G. L. & Guthrie, P. B. *Nature* **309**, 261–263 (1984).
- Ben-Ari, Y., Cherubini, E. & Krnjevic, K. *Neurosci. Lett.* **94**, 88–92 (1988).
- Kleckner, N. W. & Dingle, R. *Molec. Brain Res.* **11**, 151–159 (1991).
- Lo-Turco, J. J., Blanton, M. G. & Kriegstein, A. R. *J. Neurosci.* **11**, 792–799 (1991).
- Konnerth, A., Keller, B. U., Ballanyi, K. & Yaari, Y. *Exp. Brain Res.* **81**, 209–212 (1990).
- Sakmann, B. & Brenner, H. R. *Nature* **276**, 401–402 (1978).
- Fischbach, G. D. & Schuetze, S. M. *J. Physiol.* **303**, 125–137 (1980).
- Mishina, M. et al. *Nature* **321**, 406–411 (1986).
- Moriyoshi, K. et al. *Nature* **354**, 31–37 (1991).
- Coleman, P. A. & Miller, R. F. *J. Neurophysiol.* **61**, 218–230 (1989).
- Edwards, F. A., Konnerth, A., Sakmann, B. & Takahashi, T. *Pflügers Archs* **414**, 600–612 (1989).
- Blanton, M. G., Lo-Turco, J. L. & Kriegstein, A. R. *J. Neurosci. Meth.* **30**, 203–210 (1989).
- Trussell, L. O. & Fischbach, G. D. *Neuron* **3**, 209–350 (1989).

ACKNOWLEDGEMENTS. I thank S. Du Lac, D. M. Kullmann, B. Zalutsky, R. A. Nicoll and D. J. Perkel for comments and J. I. Korenbrot for help. This project was supported by the National Eye Institute and an NIH grant.

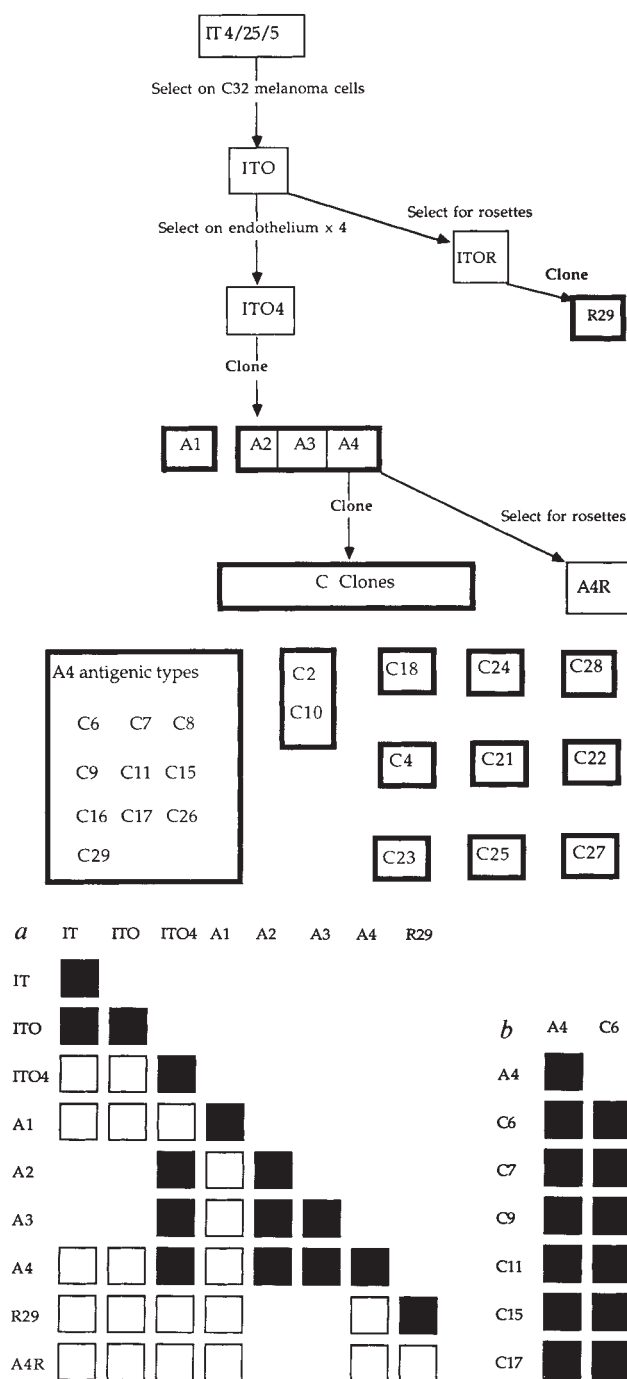


FIG. 1 Tree showing derivation of the *P. falciparum* clones from IT 4/25/5. Clones derived by micromanipulation are shown in the bold boxes. Antigenically indistinguishable clones are boxed together. R29 was cloned from ITO and A4R derived from A4 by 8 rounds of selection for rosetting. METHODS. Parasites were grown as described^{25,26} in 95% N₂, 1% O₂, 4% CO₂ using RPMI 1640 culture medium (Gibco) supplemented with 20 mM HEPES, 20 mM glucose, 5 mM glutamine and gentamycin, and 10% pooled serum from blood donors (Blood Transfusion Service, Oxford). During experiments comparing the antigenic phenotype, the cytoadherence characteristics and surface radiolabelling, clones were grown in red blood cells from the same donor and in the same pool of serum. Rosetting parasites were selected for by sedimentation through 'plasma gel' (Laboratoire Roger Bellon, Neuilly, France)²⁷. Parasites were selected for binding to C32 melanoma cells or HUVEC as previously described^{9,24}. Cloning was performed using a micromanipulator (Micro Instruments, Oxford), sterile micropipettes of 5 μ m internal diameter drawn from capillary tubes and a binocular microscope. Parasitized cells containing single trophozoites were picked from a settled monolayer by applying an aspiration pressure of 1 cm of water, transferred into a Petri dish containing supplemented RPMI 1640 and grown as above. To exclude cross-contamination, a Southern blot of *Hinf*I-digested DNA from the clones and other laboratory lines was hybridized with the probe pC4. H32 labelled with ³²P (Amersham) using a Quickprime labelling kit (Pharmacia). The autoradiograph showed that no cross-contamination had occurred (data not shown).

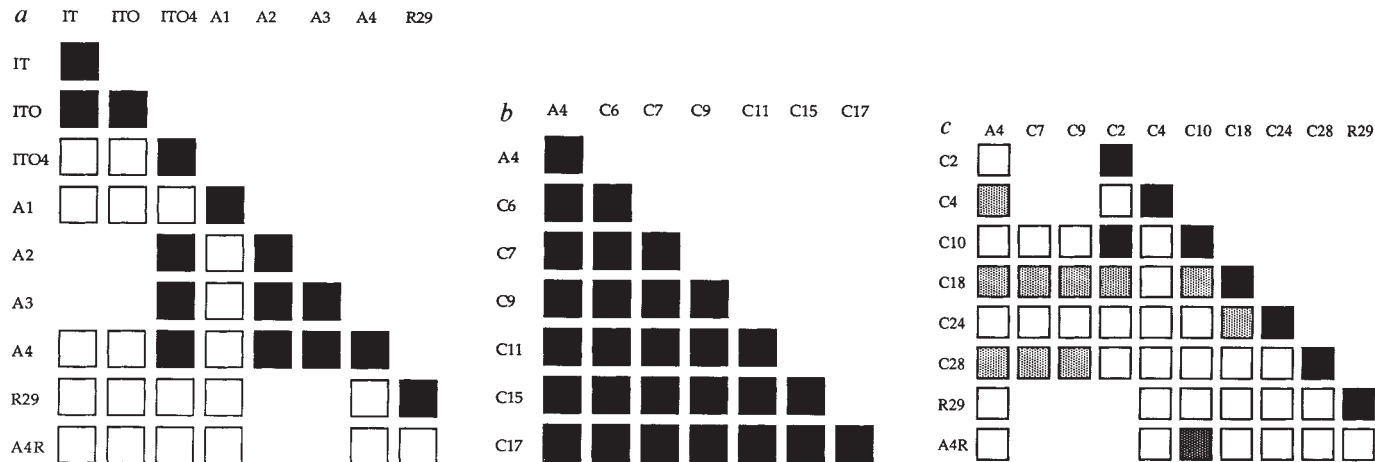


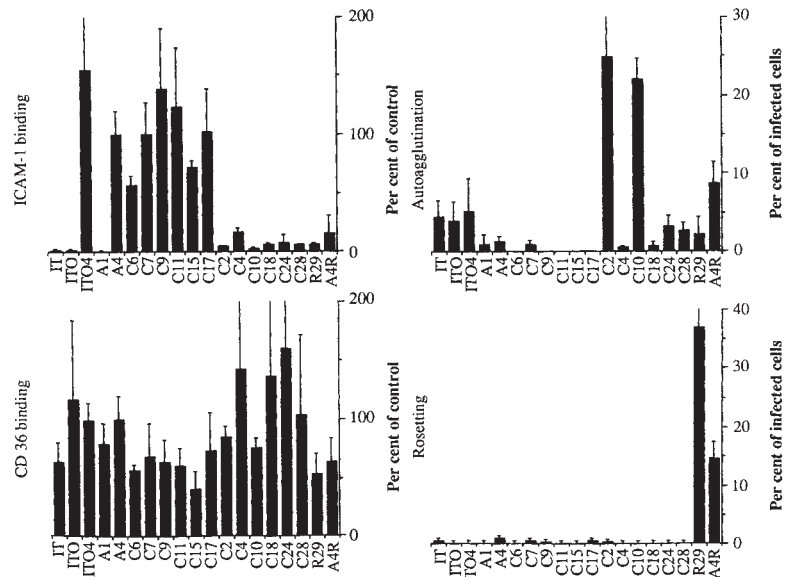
FIG. 2 Antigenic analysis of the clones. The antigenic differences between a, IT (clone IT 4/25/5), ITO, ITO4, A1-4, R29 and A4R; b, A4 and C clones with the same variant antigenic type as A4; c, A4, C7, C9, variant C clones, R29 and A4R. Key: ■, 100 to 80%; square with dark shading, 79 to 15%; lightly shaded squares, 15 to 5%; □, 5 to 0% similarity between test clones using the mixed agglutination assay. A further four C clones had the same variant type as A4, and the remaining five C clones had a different antigenic type from A4 (data not shown).

METHODS. The antigenic phenotype of the derived clones was determined by the mixed agglutination assay (C.I.N. *et al.*, submitted). In brief, the two clones to be tested were labelled independently with either ethidium bromide (200 μ g ml⁻¹) or DAPI (4, 6-diamidino-2-methylphenylindole) (20 μ g ml⁻¹) for 10 min at 37 °C, and washed three times in RPMI 1640. Labelled trophozoites (2.5 $\times$ 10⁶) were taken from each of the two cultures and resuspended in 70 μ l pooled, heat-inactivated, hyperimmune serum and 140 μ l serum-free RPMI 1640. The tube was rotated at 10 r.p.m. at room temperature for 1 h. A drop of the suspension was placed on a microscope slide and sealed with a greased coverslip. Agglutinates were examined using a Leitz epifluorescence microscope with an G365 excitation filter, an FT 395 beam splitter and an LP 420 barrier filter. Agglutinates containing 5 to 10

cells were scored for whether the agglutinate contained all red cells (stained with ethidium bromide), all blue cells (stained with DAPI) or a mixture of red and blue cells. 100 agglutinates were counted in two or more separate experiments for each pair of clones tested. The proportion of mixed-colour agglutinates is a function of the similarity of the antigenic phenotype expressed at the red blood cell surface of the two clones. The distribution of the types of agglutinate can be predicted by the binomial distribution (D.J.R., unpublished observations). If the chance of homologous agglutination is p , then for an agglutinate of size n the chance of the agglutinate being of a single colour is p^n , and the proportion of mixed-colour agglutinates is $1 - p^n$. As equal numbers of parasites from the two clones are used in the test, the ratio of chance of heterologous agglutination/chance of homologous agglutination or $1 - p/p$ is a measure of the degree of similarity between the clones. For example, assuming all agglutinates were of size 5, and the clones were antigenically indistinguishable ($p = 0.5$), the proportion of mixed agglutinates would be 0.94 and the degree of similarity would be 0.5/0.5 = 1. Similarly, mixed-colour agglutinate proportions of 0.40 and 0.20 in this test represent a degree of similarity between the two clones under test of 0.15 (15%) and 0.05 (5%), respectively.

FIG. 3 The cytoadherence phenotype of the clones. Adherence of infected red cells to *a*, purified CD36; *b*, an ICAM-1-Fc chimaera; *c*, rosetting of uninfected by infected red cells, and *d*, auto-agglutination of infected red cells. Binding to CD36 and to the ICAM-1-Fc chimaera is given as the average $\pm$ s.d. of two experiments using 4 determinations on each occasion. The percentage of infected erythrocytes rosetting is the mean $\pm$ s.d. for 500 infected cells on four occasions, and for auto-agglutination is the mean $\pm$ s.d. counting 1,000 infected cells in four determinations.

METHODS. CD36 was isolated from human platelets²⁸ and an ICAM-1-Fc chimaera prepared by transient expression in COS cells and purification on protein G-Sepharose²⁹. A solution of CD36 (3 μ l; 1 μ g ml⁻¹) or ICAM-1-Fc (3 μ l; 10 μ g ml⁻¹) were adsorbed on to a plastic dish (Falcon no. 1005) in a moist atmosphere at 4 °C overnight. The spots were aspirated and the dish washed five times with binding medium (RPMI, HEPES 20 mM, pH 6.9) before blocking with 0.5% BSA/RPMI 1640 for 4 h at 37 °C and washing 5 times in binding medium. A suspension of trophozoites (5 ml) in binding medium at 1% haematocrit and between 1 and 2% parasitaemia was placed in the dish, incubated at 37 °C, resuspending the red cells every 10 min. This was repeated five times. Unbound cells were removed from the dish by careful washing. The adherent cells were fixed with 1% glutaraldehyde, stained with Giemsa and the number of cells per high power field counted. The counts were normalized to the binding of A4 clone for both CD36 and ICAM-1 binding in each experiment. Rosetting was measured by counting the number of trophozoites rosetting two or more uninfected red cells immediately after removing a sample from the culture



flask. The percentage of trophozoites agglutinating with other infected cells (that is, auto-agglutinating) was counted after 200 μ l culture at 10% haematocrit was mixed at 10 r.p.m. for 1 h.

same antigenic type as A4 had the same cytoadherent phenotype as A4. But six C clones with a different antigenic type from A4 had a different adhesive phenotype. They showed similar binding to CD36, but greatly reduced binding to ICAM-1 compared with A4 (Fig. 3).

Neither A4 nor any C clone showed significant rosetting of uninfected erythrocytes. A rosetting line, A4R, could however be derived from A4 by selection. A4R was antigenically distinct from both A4 and all the other C clones. The variety of adhesive phenotypes that can be generated within a clone was underlined by finding a novel adhesive phenotype in two of the variant C clones. These clones, C2 and C10, showed auto-agglutination of mature infected, but not uninfected, erythrocytes in serum from donors never exposed to malaria. This phenotype has also been observed in field isolates (our unpublished observations). Rosetting and auto-agglutination are distinct phenotypes; the auto-agglutinating clones do not rosette and a freshly cloned rosetting parasite (R29) does not auto-agglutinate (Fig. 3). This distinction may be important as the relative pathological importance of the various adhesive phenotypes has not been established. Auto-agglutination *in vivo* could lead to the focal accumulation of infected erythrocytes and so contribute to the pathogenesis of severe malaria.

Clonal antigenic variation is an established mechanism of immune evasion for a variety of pathogens¹⁵⁻²⁰. It is distinguished from antigenic diversity generated by somatic mutation by the retention of a copy of each gene. This permits the subsequent appearance of previously expressed antigenic phenotypes¹⁵. It has recently been shown that a single clonal antigenic switch occurred in an antigen expressed at the surface of *P. falciparum*-infected erythrocytes²¹. To determine whether the capacity to express a previous phenotype was retained after switching, we selected a variant clone, C28, for ICAM-1 binding. After a single round of selection, the new line, C28-I, had regained ICAM-1 binding and showed increased antigenic similarity to A4, but was now antigenically distinct from C28 (Fig. 4).

Clinical immunity to malaria has been linked to antibody responses directed against the infected red-cell surface¹². Although antigenic variation *in vivo* would permit chronic infection, exposure to multiple variant antigenic types generated by

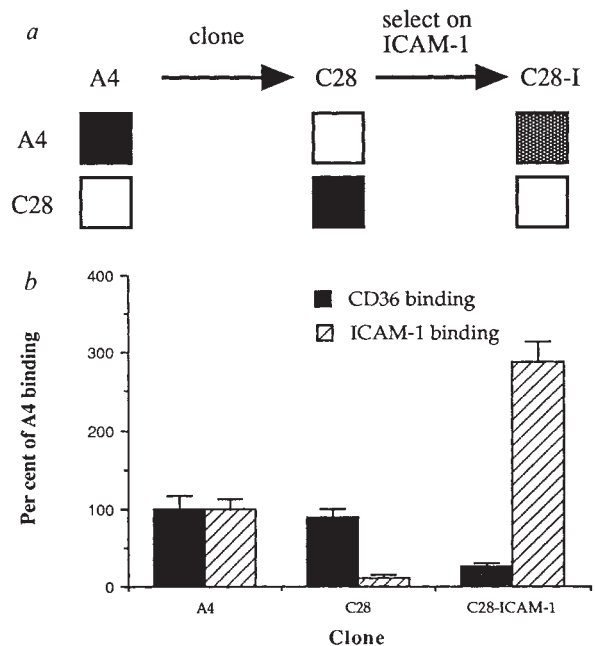


FIG. 4 The relative *a*, antigenic and *b*, adhesive phenotypes of the parental clone A4, progeny clone C28 and line C28-I derived from C28 by selection once on ICAM-1. Key: ■, 100 to 80%; shaded square, 79 to 25%; and □, 15 to 0% similarity between clones using the mixed agglutination assay. Binding to CD36 and to the ICAM-1-Fc chimaera is given as the average $\pm$ s.d. of four determinations. Results of a representative experiment are shown.

METHODS. ICAM-1-Fc chimaera (50 μ g ml⁻¹ in PBS/gentamycin) was adsorbed on to a plastic dish (Falcon no. 1005) in a moist atmosphere at 37 °C for 2 h. The solution was removed and the dish blocked with 2% BSA/PBS/gentamycin for 2 h at 37 °C and finally washed three times in binding medium. A suspension of late trophozoites (1.25 ml) purified by gelatin flotation²⁷ was allowed to settle on the dish for 1 h before the unbound parasites were washed off. Erythrocytes (50 μ l) were added to the bound cells in culture medium, and the dish placed in a gas jar at 37 °C overnight. Red cells were collected the next day and cultured as described (Fig. 1). The mixed agglutination and cytoadherence assays were done as described.

the diverse pattern of switching we have observed may prime the immune system for the rapid control of future heterologous challenges. This would explain how protective immunity can be achieved after relatively few clinical attacks of malaria^{12,22}, despite the large repertoire of variant antigenic types^{10,11}.

Changes in antigenic²³ or cytoadherence phenotypes²⁴ have been associated with a family of high-molecular-weight proteins expressed at the red-cell surface. We have found that antigenic variation is also accompanied by size changes in these proteins (data not shown). Thus the biochemical, antigenic and adhesive phenotypes are all co-modulated by clonal antigenic variation and so these properties will not be stable characteristics of an isolate, but will be determined by the interaction between the parasite genotype and the host environment. The same parasite infecting different individuals may therefore give rise to a variety

of phenotypes depending on the selective pressures exerted by the host. These pressures could include constitutive or induced levels of endothelial or red-cell surface receptor expression and the speed and effectiveness of the variant specific antibody response. Furthermore, in view of the apparent linkage between antigenic and adhesive phenotypes (Figs 2 and 3) and the large antigenic repertoire^{10,11}, additional host receptors may exist. If severe disease is associated with specific adhesive phenotypes, such as rosetting, auto-agglutination or ICAM-1 binding, characterizing the antigenic determinants of these pathogenic subsets of the total repertoire of variant antigenic types may form the basis of a realistic strategy to control severe disease. *Note added in proof:* Data linking antigenic and cytoadherence phenotypes of *P. falciparum* IT derived clones has also been obtained by Biggs *et al.*³⁰. □

Received 20 March; accepted 16 April 1992.

- MacPherson, G. G., Warrell, M. J., White, N. J., Looareesuwan, S. & Warrell, D. A. *Am. J. Path.* **119**, 385–401 (1985).
- David, P. H., Handunnetti, S. M., Leech, J. H., Gamage, P. & Mendis, K. M. *Am. J. Trop. Med. Hyg.* **38**, 289–297 (1988).
- Udomsangpetch, R. *et al. J. exp. Med.* **169**, 1835–1840 (1989).
- Carlson, J. *et al. Lancet* **336**, 1457–1460 (1990).
- Barnwell, J. W. *et al. J. clin. Invest.* **84**, 765–772 (1989).
- Ockenhouse, C. F., Tandon, N. N., Magowan, C., Jamieson, G. A. & Chulay, J. D. *Science* **243**, 1469–1471 (1989).
- Oquendo, P., Hundt, E., Lawler, J. & Seed, B. *Cell* **58**, 95–101 (1989).
- Roberts, D. D. *et al. Nature* **318**, 64–66 (1985).
- Berendt, A. R., Simmons, D. L., Tansey, J., Newbold, C. I. & Marsh, K. *Nature* **341**, 57–59 (1989).
- Marsh, K. & Howard, R. J. *Science* **231**, 150–153 (1986).
- Forsyth, K. P. *et al. Am. J. Trop. Med. Hyg.* **41**, 259–267 (1989).
- Marsh, K., Otoo, L., Hayes, R. J., Carson, D. C. & Greenwood, B. M. *Trans. R. Soc. Trop. Med.* **83**, 293–303 (1989).
- Sinnis, P. & Welles, T. E. *Genomics* **3**, 287–295 (1988).
- Turner, C. M. R. & Barry, J. D. *Parasitology* **99**, 67–75 (1989).
- Borst, P. & Greaves, D. R. *Science* **235**, 658–667 (1987).
- Brown, K. N. & Brown, I. N. *Nature* **208**, 1286–1288 (1965).

- Barnwell, J. R., Howard, R. J., Coon, H. G. & Miller, L. H. *Infect. Immun.* **40**, 985–994 (1983).
- McLean, S. A., Pearson, C. D. & Phillips, R. S. *Expl. Parasitol.* **54**, 296–302 (1986).
- Handunnetti, S. M., Mendis, K. N. & David, P. D. *J. exp. Med.* **165**, 1269–1283 (1987).
- Gilks, C. F., Walliker, D. & Newbold, C. I. *Parasite Immun.* **12**, 45–64 (1990).
- Biggs, B. A. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 9171–9174 (1991).
- Baird, J. K. *et al. Am. J. Trop. Med. Hyg.* **45**, 65–76 (1991).
- Leech, J. H., Barnwell, J. W., Miller, L. H. & Howard, R. J. *J. exp. Med.* **159**, 1567–1575 (1984).
- Magowan, C., Woolish, W., Anderson, L. & Leech, J. *J. exp. Med.* **168**, 1307–1319 (1988).
- Trager, W. & Jensen, J. B. *Science* **193**, 673–675 (1976).
- Haynes, J. D., Diggs, C. L., Hines, F. A. & Desjardins, R. E. *Nature* **263**, 767–769 (1976).
- Jensen, J. B. *Am. J. Trop. Med. Hyg.* **27**, 1274–1280 (1978).
- Tandon, N. N., Lipsky, R. H., Burgess, W. H. & Jamieson, G. A. *J. biol. Chem.* **264**, 7570–7575 (1989).
- Berendt, A. R. *et al. Cell* **68**, 71–81 (1992).
- Biggs, B. A. *et al. J. Immun.* (in the press).

ACKNOWLEDGEMENTS. We thank K. Robson for DNA samples from the FCB, 3D7 and D4 laboratory lines; R. Howard for IT 4/25/5; T. Welles for the probe pC4.H32; D. Ferguson for blood donations; B. Elford, C. Mulholland and R. Allan for help with growing clones; G. Brown, D. Weatherall, C. Bangham and S. Kyes for discussion and E. Piza for support. We thank the MRC, the Wellcome Trust, the Lister Institute for Preventive Medicine and the UNDP/World Bank/WHO special programme for Research and Training in Tropical Diseases for financial support. D.J.R. is a Medical Research Council Training Fellow and A.R.B. is a Lister Institute Research Fellow.

FK-506- and CsA-sensitive activation of the interleukin-2 promoter by calcineurin

Stephen J. O'Keefe*, Jun'ichi Tamura†‡, Randall L. Kincaid†, Michael J. Tocci* & Edward A. O'Neill*

* Department of Molecular Immunology, Merck Research Laboratories, Rahway, New Jersey 07065, USA

† Section on Immunology, Laboratory of Molecular and Cellular Neurobiology, National Institute on Alcohol Abuse and Alcoholism, Rockville, Maryland 20852, USA

ANTIGEN recognition by the T-cell receptor (TCR) initiates events including lymphokine gene transcription¹, particularly interleukin-2, that lead to T-cell activation. The immunosuppressive drugs, cyclosporin A (CsA) and FK-506, prevent T-cell proliferation by inhibiting a Ca²⁺-dependent event required for induction of interleukin-2 transcription². Complexes of FK-506 or CsA and their respective intracellular binding proteins inhibit the calmodulin-dependent protein phosphatase, calcineurin, *in vitro*³. The pharmacological relevance of this observation to immunosuppression or drug toxicity is undetermined. Calcineurin, although present in lymphocytes⁴, has not been implicated in TCR-mediated activation of lymphokine genes or in transcriptional regulation in general. Here we report that transfection of a calcineurin catalytic subunit increases the 50% inhibitory concentration (IC₅₀) of the immunosuppressants FK-506 and CsA, and that a mutant subunit acts in synergy with phorbol ester alone to activate the interleukin-2

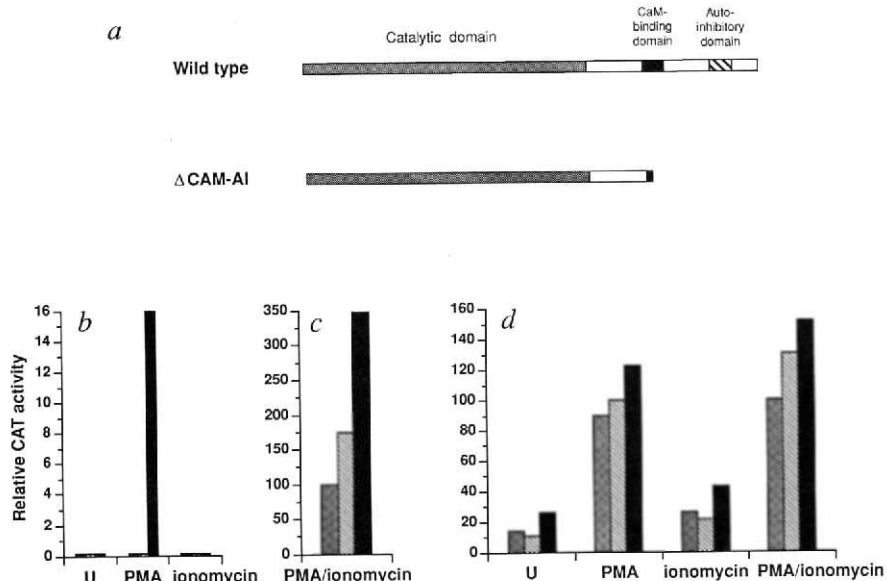
promoter in a drug-sensitive manner. These results implicate calcineurin as a component of the TCR signal transduction pathway by demonstrating its role in the drug-sensitive activation of the interleukin-2 promoter.

To investigate calcineurin's role in activation of the interleukin-2 (IL-2) promoter, we cotransfected a plasmid containing the human IL-2 promoter linked to a reporter gene along with an expression plasmid encoding either a wild type or a deletion mutant (Δ CaM-AI) of a murine calcineurin catalytic subunit (Fig. 1a) into the human T-cell line, Jurkat. Δ CaM-AI was designed to mimic proteolysed forms of the phosphatase known to have Ca²⁺-independent, constitutive phosphatase activity *in vitro*^{5,6}. Neither cotransfection of wild type nor Δ CaM-AI activated the IL-2 promoter either alone or in the presence of ionomycin (Fig. 1b). Cotransfection of wild type hyperactivated the IL-2 promoter roughly twofold in the presence of phorbol myristyl acetate (PMA; 12-O-tetradecanoylphorbol-13-acetate) and ionomycin (Fig. 1c), but had no effect with PMA alone (Fig. 1b). Thus, the combination of PMA and ionomycin is more potent in the presence of transfected calcineurin than in its absence (Fig. 1c), suggesting that calcineurin phosphatase activity limits the level of transcription from the IL-2 promoter. As expected for a constitutively activated component of an essential Ca²⁺-dependent signalling pathway, Δ CaM-AI acted in synergy with PMA, bypassing the Ca²⁺ requirement and inducing >150-fold activation of the IL-2 promoter (Fig. 1b). Cotransfection of Δ CaM-AI also increased promoter activity in the presence of both mitogens 3.5-fold (Fig. 1c). The dependence on ionomycin for maximal promoter activity may reflect activation of endogenous calcineurin and/or activation of Ca²⁺-dependent, calcineurin-independent pathways.

To demonstrate that the observed results were directly attributable to IL-2 promoter activation, we investigated Δ CaM-AI's effect on a reporter plasmid in which the cytomegalovirus immediate-early (CMV-IE) promoter directs chloramphenicol

† Present address: Third Department of Internal Medicine, Gunma University Medical School, Maebashi, Gunma 371, Japan.

FIG. 1 A deletion mutant of murine calcineurin acts in synergy with PMA in FK-506-sensitive activation of the IL-2 promoter. **a**, Schematic representation of wild-type calcineurin catalytic subunit and the deletion mutant, Δ CaM-AI. The putative autoinhibitory¹³ and CaM-binding domains¹⁴ are indicated. Wild type was constructed by inserting CN α 4¹⁵, a complementary DNA clone encoding amino acids 1 to 521, into the *Eco*RI site of the eukaryotic expression vector pCDL-SR α 296¹⁶. The deletion mutant was constructed by polymerase chain reaction (PCR) amplification of a CN α -4 template and insertion of a fragment encoding amino acids 1 to 398 into the *Sma*I site of pUC18¹⁷. Consequently, the 5' and 3' untranslated sequences present in CN α 4 were also deleted. The clone was subsequently inserted into the *Eco*RI site of pCDL-SR α 296. The plasmids are designated pSR α -calcineurin and pSR α - Δ CaM-AI, respectively. The wild type construct encodes a murine catalytic subunit 99% identical to the corresponding human subunit¹⁵. Δ CaM-AI lacks functional CaM-binding¹⁴ and autoinhibitory¹³ domains. This truncation was designed to mimic proteolysed forms of the phosphatase known to have Ca²⁺-independent, constitutive phosphatase activity *in vitro*^{5,6}. **b**, Δ CaM-AI transactivates the IL-2 promoter in synergy with PMA. The reporter, p(An)IL2.CAT, contains the human IL-2 promoter (base pairs -448 to +43) directing transcription of the CAT gene and includes an SV40 polyadenylation site between vector sequences and -448 of the IL-2 promoter to prevent vector-dependent transcription from contributing to the observed CAT activity. The p(An)IL2.CAT was cotransfected into Jurkat cells together with the effectors, pCDL-SR α 296 (vector, $\square$), pSR α -calcineurin ($\square$), or pSR α - Δ CaM-AI ($\blacksquare$) and the transfection efficiency control, pCMV.Gal which contains the CMV-IE promoter directing transcription of the bacterial β -galactosidase (β -gal) gene. The transfected cells were activated as indicated 16 h after electroporation⁷, incubated for another 16 h, collected and assayed. The data shown are an average of three or four independent experiments. **c**, Calcineurin hyperactivates the IL-2 promoter with PMA and ionomycin. Transfections were done as described above. **d**, Δ CaM-AI has no significant effect on the PMA-dependent stimulation of the CMV-IE promoter. The reporter, pCMV.CAT, contains the CMV-IE promoter (base pairs -760 to +4) directing transcription of the CAT gene. Transfections were done as



described above. The data shown are an average of two independent experiments.

METHODS. Jurkat cells were passaged as described⁷. Cells were grown to a density of $4-6 \times 10^5 \text{ ml}^{-1}$. Transfection was by electroporation (10^7 cells in $300 \mu\text{l}$ RPMI medium, 10% FCS, room temperature, $960 \mu\text{F}$, 250 V , 0.4 cm cuvette, BioRad Gene Pulser) using $2 \mu\text{g}$ reporter plasmid, $15 \mu\text{g}$ of effector plasmid and $0.5-2 \mu\text{g}$ of the transfection efficiency control pCMV.Gal. Before activation or drug treatment each culture of transfected cells was divided into the appropriate number of aliquots. Cells were collected 12-16 h after treatment, lysed, and assayed for CAT¹⁸ and β -gal¹⁹ activities. The β -gal activity in PMA-treated cells was unaffected by the presence of the transfected calcineurin as demonstrated by addition of FK-506 to PMA-treated cells (data not shown). Relative CAT activity was adjusted for transfection efficiency as dictated by β -gal activity. For each promoter, the CAT activity induced by the indicated treatment is shown relative to the PMA/ionomycin-induced CAT activity of the vector control, which was set to 100.

acetyltransferase (CAT) gene transcription. This promoter has a measurable basal transcription activity and is stimulated by PMA in an FK-506-resistant manner⁷. Δ CaM-AI increases the basal activity of the CMV-IE promoter slightly (<twofold), but has virtually no effect on the PMA-stimulated activity (Fig. 1d). Thus, the effect of Δ CaM-AI is not due to stabilization of CAT message or activation of CAT enzyme.

If calcineurin is the FK-506-sensitive component in TCR signal transduction, then FK-506 should inhibit PMA/ Δ CaM-AI-mediated activation of the IL-2 promoter. In fact, when FK-506 was added (to 10 ng ml^{-1}) immediately before PMA,

activation of the IL-2 promoter was completely inhibited (Fig. 2a). Rapamycin, a macrolide that is structurally related to FK-506 but that does not inhibit calcineurin *in vitro*³, had no effect at 100 ng ml^{-1} (Fig. 2a). Neither compound affected PMA-dependent activation of the CMV-IE promoter (Fig. 2b). Inhibition of Δ CaM-AI by FK-506 in the absence of ionomycin suggests that deletion of the autoinhibitory and calmodulin (CaM)-binding domains greatly reduces the Ca²⁺ requirement for binding of calcineurin to FK-506/FKBP³. Experiments in which trypsinized calcineurin is inhibited by CsA/cyclophilin A in the presence of 1 mM EGTA support this conclusion⁸.

FIG. 2 FK-506 inhibits Δ CaM-AI-dependent activation of transcription. The plasmids p(An)IL2.CAT (**a**) and pCMV.CAT (**b**) were cotransfected with pSR α - Δ CaM-AI and the transfected cells were manipulated as described in Fig. 1. FK-506 (10 ng ml^{-1} , $\square$) or rapamycin (100 ng ml^{-1} , $\square$) was added immediately before addition of PMA ($\blacksquare$) as indicated. Relative CAT activity is calculated as described in Fig. 1. The results presented in (**a**) are the average of four independent experiments and those in (**b**) of two independent experiments.

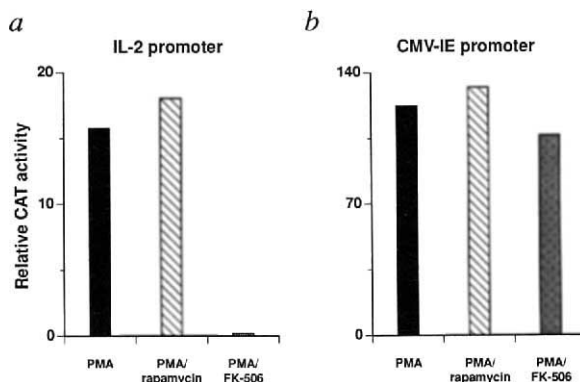
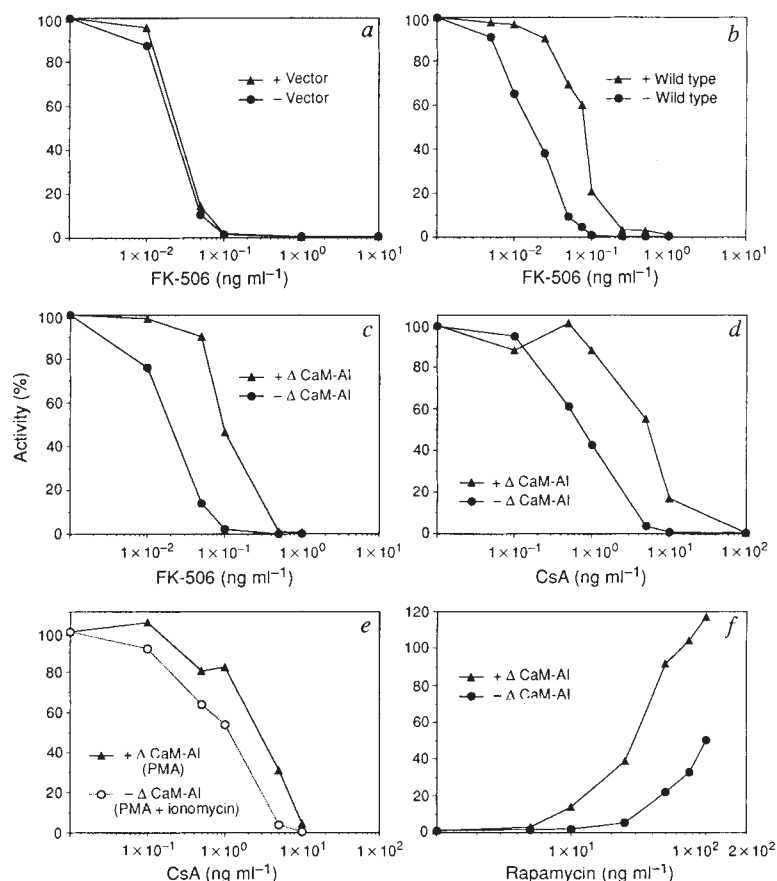


FIG. 3 Expression of the Δ CaM-AI mutant increases the IC_{50} of FK-506 and CsA and decreases the ED_{50} of rapamycin required to antagonize FK-506. All drug titrations were done with cell cultures that had been stimulated with the combination of PMA and ionomycin except as indicated. CAT ($\blacktriangle$) and β -gal ($\bullet$) activities are plotted as per cent of activity in lysates of transfected cells that were not treated with any immunosuppressive drug. **a**, Effect of vector on IC_{50} of FK-506. **b**, Effect of wild type on IC_{50} of FK-506. **c**, Effect of Δ CaM-AI on IC_{50} of FK-506. **d**, Effect of Δ CaM-AI on IC_{50} of CsA. **e**, Effect of Δ CaM-AI on IC_{50} of CsA in the absence of ionomycin. The β -gal titration ($\circ$) was determined from a parallel culture treated with the combination: PMA and ionomycin. **f**, Effect of Δ CaM-AI on ED_{50} of rapamycin.

METHODS. In each experiment Jurkat cells were cotransfected (as in Fig. 1) with p(An)IL2.CAT and either pcDL-SR α 296 (**a**), pSR α -calcineurin (**b**), or pSR α - Δ CaM-AI (**c-f**), except the concentration of the effector plasmid was 30 μ g per transfection and the pCMV.Gal was omitted. As an internal control for efficacy of the immunosuppressive drugs, Jurkat cells were transfected independently with 30 μ g of a plasmid containing a derivative of the human IL-2 promoter directing transcription of the β -gal gene (pIL-2.Gal) and mixed with the CAT/effector cotransfected cells. After 24 h the cells were divided into aliquots and activated with PMA and ionomycin in the presence of the indicated concentrations of FK-506 (**a-c**), CsA (**d** and **e**), or 0.5 ng ml $^{-1}$ FK-506 and the indicated concentrations of rapamycin (**f**). The cells were collected after 16 h and assayed as in Fig. 1.



Having postulated that hyperactivation of the IL-2 promoter by calcineurin expression plasmids is due to an increase in calcineurin concentration, we would predict that the IC_{50} of FK-506 should increase after transfection. The expression vector alone had no effect (Fig. 3a), but the IC_{50} of FK-506 increased 4–4.5-fold in the presence of both wild type and Δ CaM-AI. In representative experiments, the IC_{50} increased from 0.02 ng ml $^{-1}$ to 0.08 ng ml $^{-1}$ for wild type (Fig. 3b) and to 0.09 ng ml $^{-1}$ for Δ CaM-AI (Fig. 3c). Similarly, Δ CaM-AI increases the IC_{50} of CsA over six-fold, from 0.8 ng ml $^{-1}$ to 5 ng ml $^{-1}$ in the depicted titration (Fig. 3d). The wild-type calcineurin cannot bind FK-506/FKBP in the absence of Ca^{2+} *in vitro*³; therefore, if the concentration of endogenous calcineurin is significant relative to that of the transfected calcineurin, the IC_{50} of CsA or FK-506 with Δ CaM-AI and PMA should be lower than that with Δ CaM-AI, PMA and ionomycin. As predicted, the IC_{50} of CsA, in the absence of ionomycin, decreases. In a typical experiment, the IC_{50} is reduced from 5 ng ml $^{-1}$ to 3 ng ml $^{-1}$ (Fig. 3e).

Rapamycin antagonizes the immunosuppressive effects of FK-506⁹, does not inhibit calcineurin *in vitro*³, and has no effect on

Δ CaM-AI-dependent stimulation of the IL-2 promoter (Fig. 2a). To relate the FK-506 inhibition described above to known pharmacological characteristics of FK-506, we examined the ability of rapamycin to antagonize FK-506 in the presence of Δ CaM-AI. Typically, the 50% effective dose (ED_{50}) of rapamycin decreased three-fold in the presence of Δ CaM-AI (Fig. 3f), consistent with the hypothesis that the change in IC_{50} for each immunosuppressant results from increasing the concentration of calcineurin.

Several genes encoding catalytic subunits of calcineurin have been identified¹⁰. This study indicates that a truncated form of the type 1 (or α) gene acts in synergy with PMA to activate the IL-2 promoter. Our results provide the first biological evidence that calcineurin is an effector of Ca^{2+} -dependent transcriptional activation and is the FK-506-sensitive component of the TCR signal transduction pathway. In addition to IL-2 promoter activation in T cells, calcineurin may play a role in other Ca^{2+} -dependent, FK-506-sensitive events such as neutrophil degranulation¹¹, inhibition of Ly-6E gene transcription¹² and various CsA/FK-506 toxicities in other organs. $\square$

Received 18 March; accepted 12 June 1992.

- Crabtree, G. R. *Science* **243**, 355–361 (1989).
- Sigal, N. H. & Dumont, F. J. in *A. Rev. Immun.* 519–560 (Annual Reviews, Palo Alto, 1992).
- Liu, J. et al. *Cell* **66**, 807–815 (1991).
- Kincaid, R. L., Takayama, H., Billingsley, M. L. & Sitkovsky, M. V. *Nature* **330**, 176–178 (1987).
- Hubbard, M. J. & Klee, C. B. *Biochemistry* **28**, 1868–1874 (1989).
- Kincaid, R. L., Martensen, T. M. & Vaughn, M. *Biochem. biophys. Res. Commun.* **140**, 320–328 (1986).
- Banerji, S. S., Parsons, J. N. & Tocci, M. J. *Molec. cell. Biol.* **11**, 4074–4087 (1991).
- Swanson, S. K.-H. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3741–3745 (1992).
- Dumont, F. J. et al. *J. Immun.* **144**, 1418–1424 (1990).
- Kincaid, R. L., Higuchi, S., Tamura, J., Giri, P. R. & Martensen, T. M. *Adv. Prot. Phosphatases* **6**, 73–98 (1991).

- Forrest, M. J., Jewell, M. E., Koo, G. C. & Sigal, N. H. *Biochem. Pharmac.* **42**, 1221–1228 (1991).
- Altmeyer, A. et al. *Int. J. Immunopharm.* **13**, 1187–1189 (1991).
- Hashimoto, Y., Perrino, B. A. & Soderling, T. R. *J. Biol. Chem.* **265**, 1924–1927 (1990).
- Kincaid, R. L., Nightingale, M. S. & Martin, B. M. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8983–8987 (1988).
- Kincaid, R. L. et al. *J. Biol. Chem.* **265**, 11312–11319 (1990).
- Takebe, Y. et al. *Molec. cell. Biol.* **8**, 466–472 (1988).
- Tamura, J. & Kincaid, R. L. *FASEB J.* **4**, 2236 (1990).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
- Hollon, T. & Yoshimura, F. K. *Analyt. Biochem.* **182**, 411–418 (1989).

ACKNOWLEDGEMENTS. We thank R. Boris for FK-506 and rapamycin, P. Durette for CsA, N. Arai (DNAX, Palo Alto) for the plasmid, pcDL-SR α 296, J. Siekierka for the Jurkat cells, B. Frantz, E. Nordby and J. N. Parsons for technical assistance and A. Williamson, N. Sigal, G. Wiederrecht, J. Siekierka, B. Frantz and F. Dumont for reading the manuscript and helpful discussions.

Identification of calcineurin as a key signalling enzyme in T-lymphocyte activation

Neil A. Clipstone & Gerald R. Crabtree

Beckman Center for Molecular and Genetic Medicine,
Howard Hughes Medical Institute, Stanford University School of Medicine,
Stanford, California 94305, USA

THE immunosuppressive drugs cyclosporin A (CsA) and FK506 both interfere with a Ca^{2+} -sensitive T-cell signal transduction pathway¹⁻⁴, thereby preventing the activation of specific transcription factors (such as NF-AT and NF-IL2A)^{1,5-7} involved in lymphokine gene expression. CsA and FK506 seem to act by interaction with their cognate intracellular receptors⁸⁻¹⁰, cyclophilin and FKBP, respectively (see ref. 11 for review). The Ca^{2+} /calmodulin-regulated phosphatase calcineurin is a major target of drug-isomerase complexes *in vitro*¹². We have therefore tested the hypothesis that this interaction is responsible for the *in vivo* effects of CsA/FK506. We report here that overexpression of calcineurin in Jurkat cells renders them more resistant to the effects of CsA and FK506 and augments both NFAT- and NFIL2A-dependent transcription. These results identify calcineurin as a key enzyme in the T-cell signal transduction cascade and provide biological evidence to support the notion that the interaction of drug-isomerase complexes with calcineurin underlies the molecular basis of CsA/FK506-mediated immunosuppression.

To determine directly the potential *in vivo* biological relevance of the drug-isomerase-calcineurin interaction observed *in vitro*¹², we examined the effects of calcineurin (CN) overexpression on the sensitivity of T-cell activation to CsA and FK506. Because calcineurin exists in multiple isoforms¹³⁻¹⁵, we first determined whether the murine CN- α_1 isoform used in this study (CN α_4 in ref. 13) could interact with drug-isomerase complexes. To this end, expression vectors encoding the murine catalytic CN- α_1 and regulatory CN-B subunits (pBJ5-CNA and pBJ5-CNB, respectively) were cotransfected into TAG Jurkat cells. The murine CN- α_1 isoform was readily detected by immunoblotting in extracts from transfected cells but not mock transfected cells and was found to interact specifically with a glutathione S-transferase/FKBP-12 fusion protein in an FK506-dependent fashion (Fig. 1).

The *cis*-acting elements in the promoter of the interleukin-2 (IL-2) gene that are the targets for the action of CsA and FK506 have been identified^{1,5,6}. The transcription factors that bind to two of these sites, namely NF-AT and NF-IL2A, are critical for IL-2 gene expression¹⁶ and their function is inhibited by FK506 and CsA in a dose-dependent fashion^{1,5}. We therefore used plasmids containing multiple copies of binding sites for either NF-AT or NF-IL2A located upstream of the minimal IL-2 promoter and the secreted alkaline phosphatase reporter gene¹⁷ (NFAT-SX and NFIL2A-SX, respectively), to monitor the effects of CsA and FK506 on TAG Jurkat cells. Transfection of both pBJ5-CNA and pBJ5-CNB resulted in a large increase in the resistance of TAG cells to the effects of FK506 relative to cells transfected with the expression vector pBJ5 alone (Fig. 2a, b). Similarly, overexpression of calcineurin also made cells more resistant to the effects of CsA (Fig. 2c, d). This effect was observed whether cells were stimulated either with ionomycin and phorbol myristyl acetate (PMA; 12-O-tetradecanoyl-phorbol-13-acetate) (Fig. 2a-d) or anti-CD3 monoclonal antibody and PMA (data not shown). Overexpression of the human protein phosphatase 2A (PP2A) gene, which is highly related to calcineurin in its catalytic domain¹³⁻¹⁵, had no significant effect on the FK506-mediated inhibition of NFAT-SX activity (Fig. 2a, open squares), indicating that the ability to make cells more resistant to the immunosuppressive effects of CsA and FK506 seems specific to calcineurin.

How does overexpression of calcineurin make Jurkat cells resistant to the effects of the immunosuppressive drugs CsA and FK506? One possibility is that calcineurin is an essential enzyme in the T-cell signalling cascade. Consequently, overexpression of the enzyme should result in an excess of calcineurin relative to the concentration of the inhibitory drug-isomerase complex, therefore overcoming the inhibitory effect of the drug and allowing signalling to proceed normally. Alternatively, it is possible that overexpression of calcineurin acts indirectly, in that it does not actually play a role in T-cell signalling, but rather binds and sequesters the drug-isomerase complex preventing it from interacting with its relevant intracellular target.

To distinguish between these alternative possibilities we examined the effects of calcineurin overexpression on the signalling requirements for NFAT- and NFIL2A-dependent transcription in transfected TAG Jurkat cells. Optimal expression of both NFAT- and NFIL2A-dependent transcription requires two stimuli, namely phorbol ester and calcium ionophore^{1,18}, that activate protein kinase C and increase the concentration of intracellular calcium, respectively. As calcineurin is a calcium-regulated phosphatase¹⁹ and thus could partly mediate the effects of the calcium signal during T-cell activation, we investigated whether overexpression of calcineurin could affect the requirement for an increase in intracellular calcium. Thus, TAG

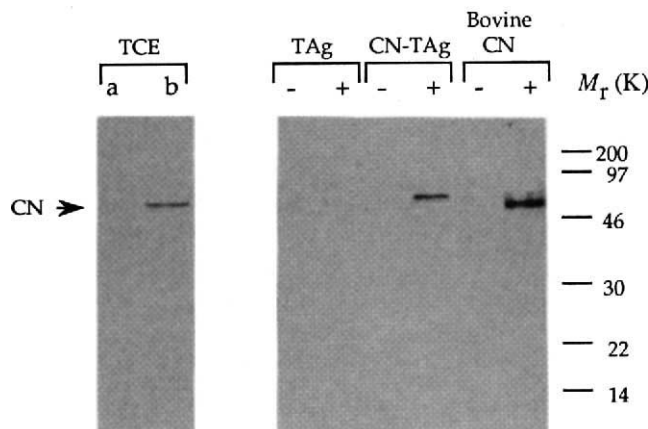
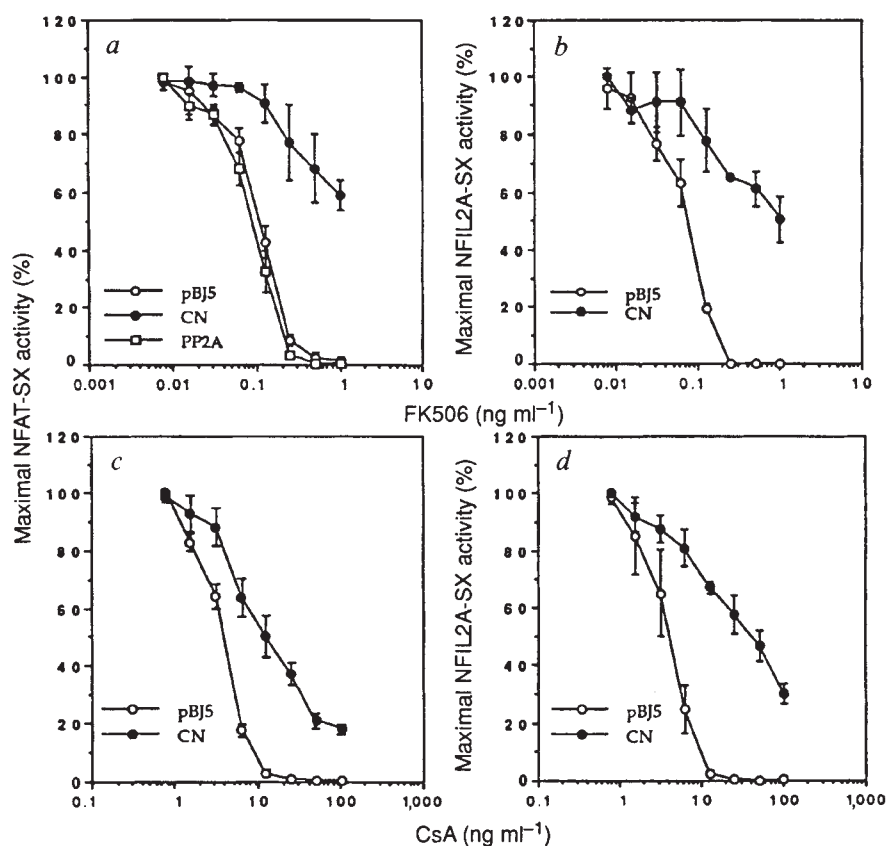


FIG. 1 The murine CN α_1 isoform interacts specifically with FKBP-FK506 complexes. Left, Immunoblot analysis using goat anti-CN antibodies of total cell extracts (TCE) from mock transfected (lane a) or CN-transfected TAG Jurkat cells (lane b). Right, Immunoblot analysis using goat anti-CN antibodies of proteins specifically bound by GST-FKBP-12 fusion proteins in the absence (-) or presence (+) of FK506 from mock transfected TAG Jurkat cells (TAG), CN-transfected TAG Jurkat cells (CN-TAG) or as a positive control purified bovine brain CN (Sigma; Bovine CN). The minor cross-reacting band observed in all lanes, both in the presence and absence of FK506, does not comigrate with calcineurin.

METHODS. TAG Jurkat cells, a derivative of the human T-cell leukaemia line Jurkat stably transfected with the SV40 large T antigen (J. Northrop and G.R.C., unpublished results), were maintained in RPMI 1640 medium containing 10% (v/v) fetal calf serum. Because the pBJ5 expression vector contains an SV40 origin of replication, transfection into TAG Jurkat cells results in high-level gene expression. TAG Jurkat cells were transiently transfected by electroporation with 2 μg pBJ5-CNA and 2 μg pBJ5-CNB or 4 μg pBJ5¹. Cells were lysed 48 h after transfection in lysis buffer (25 mM HEPES pH 7.2, 100 mM NaCl, 2 mM MgCl_2 , 2 mM CaCl_2 , 1 mM PMSF, 10 $\mu\text{g ml}^{-1}$ leupeptin and 0.5% Triton X-100). Immobilized glutathione S-transferase/FKBP-12 fusion protein complexes made in the presence or absence of FK506 were mixed with extracts derived from 2.5×10^7 transfected cells or with 4 μg CN (Sigma) and 16 μg calmodulin (Sigma) as a positive control. After incubation for 3 h at 4 $^\circ\text{C}$, GST-FKBP-bound complexes were washed five times with lysis buffer and analysed by immunoblotting with goat anti-CN antibodies²². Immunoreactive bands were visualized using horseradish peroxidase-coupled rabbit anti-goat immunoglobulin and the ECL system (Amersham). Samples corresponding to total cell extracts were derived from 5×10^5 cells. The arrow indicates CN and the position of molecular standards are shown to the right.

FIG. 2 Overexpression of murine calcineurin renders TAG Jurkat cells more resistant to CsA and FK506. TAG Jurkat cells were cotransfected either with NFAT-SX (a, c) or NFIL2A-SX (b, d) together with either pBJ5-CNA and pBJ5-CNB (filled circles), pBJ5 alone (open circles) or pBJ5-PP2A (open squares). Transfected cells were stimulated with ionomycin and PMA in the presence of various concentrations of FK506 (a, b) or CsA (c, d) and assayed for secreted alkaline phosphatase activity.

METHODS. TAG Jurkat cells (10^7) were transiently transfected, as described previously², with 2 μ g of the indicated plasmid, the amount of DNA was kept constant by the addition of pBJ5. Twenty-four h after transfection, 2×10^5 cells were aliquoted into wells of a 96-well plate (Costar) and cultured in a final volume of 200 μ l growth medium containing 0.5 μ M ionomycin 10 ng ml^{-1} PMA and various concentrations of either CsA or FK506. After a 20 h stimulation period, cell suspensions were heated to 67 °C for 60 min to inactivate endogenous cellular phosphatases and assayed for heat-stable secreted alkaline phosphatase activity. Aliquots of cell suspension (100 μ l) were mixed with 100 μ l assay buffer (2 M diethanolamine bicarbonate pH 10, 19 mM L-homoarginine and 1.2 mM methylumbelliferyl phosphate) and after incubation at 37 °C for 12 h, fluorescence at 460 nm was measured using 355 nm excitation with a Titertek Fluoroscanner II. Control experiments revealed that TAG cells transfected with pBJ5-CNA and pBJ5-CNB in the absence of reporter constructs did not exhibit phosphatase activity above background levels. Secreted alkaline phosphatase (SEAP) activity was determined in duplicate for each experimental point in the dose-response series. Nonstimulated levels of SEAP activity were <5 arbitrary units (a.u.), maximal stimulated levels were typically in the range 400–600 a.u. for OCT-SX and 600–1000 for NFAT-SX. The results are expressed as per cent of maximal reporter construct activity and represent the mean of values



determined from three independent transfections. Error bars represent standard deviations. The data is representative of at least four independent experiments and was repeated with different plasmid preparations. Qualitatively identical results were obtained with chloramphenicol acetyltransferase (CAT)-based reporter constructs.

FIG. 3 Overexpression of murine calcineurin augments NFAT- and NFIL2A-dependent transcription. TAG Jurkat cells were cotransfected with either NFAT-SX (a) or NFIL2A-SX (b) together with either pBJ5-CNA and pBJ5-CNB (filled circles), pBJ5 alone (open circles) or pBJ5-PP2A (open squares). Transfected cells stimulated with a constant concentration of PMA and various concentrations of ionomycin were assayed for secreted alkaline phosphatase activity.

METHODS. TAG Jurkat cells were transiently transfected with the indicated plasmids, stimulated with 10 ng ml^{-1} PMA and various concentrations of ionomycin and then assayed for secreted alkaline phosphatase activity as described in the legend to Fig. 2. SEAP activity was determined in duplicate for each experimental point in the dose-response series. The results are expressed as per cent of maximal reporter construct activity and represent the mean of values determined from three independent transfections. Error bars represent standard deviations. The data is representative of at least four independent experiments and was repeated with different plasmid preparations. Qualitatively identical results were obtained with CAT-based reporter constructs.

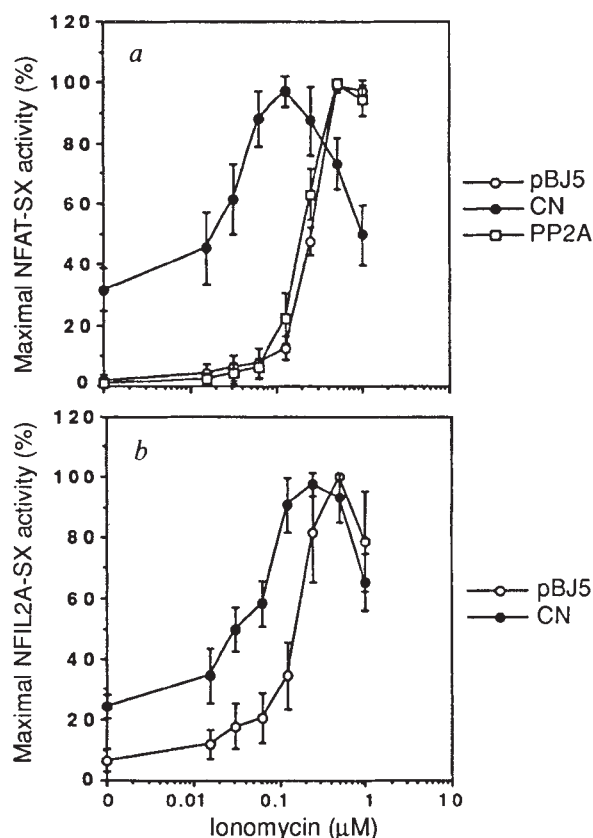
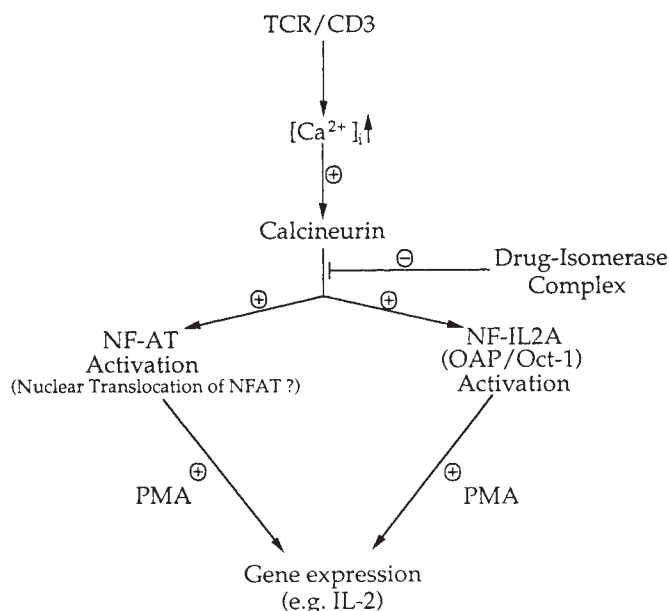


FIG. 4 A schematic model to illustrate the proposed central role of calcineurin in T-cell activation. We do not mean to infer from this model, however, that calcineurin is involved in all Ca^{2+} -dependent transcriptional events in T cells, because the transcription of certain genes, although requiring an increase in intracellular calcium, remain insensitive to the effects of CsA and FK506^{1,4}.



Jurkat cells were cotransfected with pBJ5-CNA and pBJ5-CNB or as a control pBJ5 alone, together with either NFAT-SX or NFIL2A-SX, then stimulated with a constant concentration of PMA (10 ng ml^{-1}) and various concentrations of the calcium ionophore ionomycin. Transfection of murine calcineurin augments both NFAT- and NFIL2A-dependent transcription in response to elevated intracellular calcium levels (Fig. 3). Note that high levels of ionomycin actually inhibit reporter gene activity. This effect was also augmented by overexpression of calcineurin and may reflect the deleterious consequences for the T cell of a supraoptimal Ca^{2+} signal. Significant reporter gene activity was detected in calcineurin-transfected cells at sub-optimal concentrations of ionomycin that elicited little or no activity in control cells. Moreover, in the absence of ionomycin, transfected murine calcineurin can act in synergy with PMA to stimulate ~30% of maximal NFAT-SX and NFIL2A-SX activity. This apparent Ca^{2+} /calmodulin-independent effect of calcineurin is probably a direct consequence of overexpression of the CN-A and CN-B subunits, that together have significant *in vitro* phosphatase activity in the absence of calmodulin²⁰. Confirmation of the specificity of the effect of calcineurin was provided by two independent observations. First, cotransfection of the human PP2A gene had no effect on the ionomycin dose-response curves for NFAT-SX (Fig. 3a, open squares), and second, overexpression of calcineurin did not affect the PMA inducibility of the CsA/FK506-insensitive AP-1 reporter gene (data not shown).

The data presented here identify calcineurin as a key rate-limiting enzyme in the T-cell signal transduction cascade and provide biological evidence to support the notion that the interaction of drug-isomerase complexes with calcineurin underlies the molecular basis of CsA and FK506 action. The recent finding of a strong correlation between the immunosuppressive properties of a range of CsA/FK506 analogues and their ability to bind calcineurin²¹, provides further support for this model. We have shown that overexpression of murine calcineurin renders cells more resistant to the effects of CsA and FK506. *A priori* these findings suggest that the primary determinant of cellular sensitivity to these immunosuppressive drugs is in fact the intracellular concentration of calcineurin. T lymphocytes express low levels of calcineurin^{13,22}, which may help to explain the exquisite sensitivity of this cell lineage to CsA and FK506. We have also demonstrated that overexpression of calcineurin can act in synergy with PMA-mediated signals to stimulate NFAT- and NFIL2A-dependent transcription, which provides

evidence linking calcineurin to the activation of two transcription factor complexes that play essential roles in IL-2 gene expression¹⁶ and possibly the expression of other lymphokine genes⁶. This establishes calcineurin as a nodal point in the transduction of signals from the early biochemical events that occur at the T-cell membrane to gene regulatory events in the nucleus (Fig. 4), and thereby provides a further explanation for the profound effects of CsA and FK506 on T-cell activation.

How does calcineurin regulate the activity of NF-AT and NF-IL2A? CsA and FK506 both prevent the calcium-mediated translocation of a component of the NF-AT complex from the cytoplasm to the nucleus⁷. Thus, calcineurin could potentially directly regulate the nuclear translocation of the NF-AT cytoplasmic component. In this regard, the yeast transcription factor SWI5 presents a well characterized precedent, because its cell cycle-dependent nuclear translocation is regulated by the dephosphorylation of serine residues located in and close to the nuclear translocation signal²³. Alternatively, calcineurin may regulate NF-AT function indirectly. The relevant calcineurin substrates need to be identified to resolve these alternatives. □

Received 6 April; accepted 12 June 1992.

1. Mattila, P. S. *et al.* *EMBO J.* **9**, 4425-4433 (1990).
2. Lin, C. S., Boltz, R. C., Siekierka, J. J. & Sigal, N. H. *Cell Immun.* **133**, 269-284 (1991).
3. Kay, J. E., Doe, S. E. A. & Benzie, C. R. *Cell Immun.* **124**, 175-181 (1989).
4. Gunter, K. C., Irving, S. G., Zipfel, P. F., Siebenlist, U. & Kelly, K. J. *Immun.* **142**, 3286-3291 (1989).
5. Emmel, E. A. *et al.* *Science* **246**, 1617-1620 (1989).
6. Randak, C., Brabletz, T., Hergenrother, M., Sobotta, I. & Serfling, E. *EMBO J.* **9**, 2529-2536 (1990).
7. Flanagan, W. F., Corthesy, B., Bram, R. J. & Crabtree, G. R. *Nature* **352**, 803-807 (1991).
8. Bierer, B. E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 9231-9235 (1990).
9. Handschumacher, R. E., Harding, M. W., Rice, J. & Druggie, R. J. *Science* **226**, 544-547 (1984).
10. Tropschug, M., Barthelmess, I. B. & Neupert, W. *Nature* **342**, 953-955 (1989).
11. Schreiber, S. L. *Science* **251**, 283-287 (1991).
12. Liu, J. *et al.* *Cell* **66**, 807-815 (1991).
13. Kincaid, R. L. *et al.* *J. Biol. Chem.* **265**, 11312-11319 (1990).
14. Guerini, D. & Klee, C. B. *Proc. natn. Acad. Sci. U.S.A.* **86**, 9183-9187 (1989).
15. McPartlin, A. E., Barker, H. M. & Cohen, P. T. W. *Biochem. biophys. Acta* **1088**, 308-310 (1991).
16. Durand, D. B. *et al.* *Molec. cell. Biol.* **8**, 1715-1724 (1988).
17. Berger, J., Hauber, J., Hauber, R., Geiger, R. & Cullen, B. R. *Gene* **66**, 1-10 (1988).
18. Flanagan, W. M. & Crabtree, G. R. *J. Biol. Chem.* **267**, 399-406 (1992).
19. Klee, C. B., Draetta, G. F. & Hubbard, M. J. in *Advances in Enzymology and Related Areas of Molecular Biology* (ed. Meister, A.) 149-209 (Wiley, New York, 1988).
20. Merat, D. L., Hu, Z. Y., Carter, T. E. & Cheung, W. Y. *J. Biol. Chem.* **260**, 11053-11059 (1985).
21. Lui, J. *et al.* *Biochemistry* **31**, 3896-3901 (1992).
22. Kincaid, R. L., Takayama, H., Billingsley, M. L. & Sitkovsky, M. V. *Nature* **330**, 176-178 (1987).
23. Moll, T., Tebb, G., Surana, U., Robitsch, H. & Nasmyth, K. *Cell* **66**, 743-758 (1991).

ACKNOWLEDGEMENTS. We thank R. Kincaid for anti-CN antibodies and cDNAs encoding the murine CN-A and CN-B subunits, R. Bram for plasmids NFAT-SX and NFIL2A-SX and discussions, P. Kao for GST-FKBP-12, J. Northrop for TAg Jurkat cells and M. Flanagan and D. Spencer for reading the manuscript. This work was supported by the Howard Hughes Medical Institute and the NIH. N.A.C. is a fellow of the Anna Fuller Fund.

Alterations in a yeast protein resembling HIV Tat-binding protein relieve requirement for an acidic activation domain in GAL4

Jonathan C. Swaffield, Jacqueline F. Bromberg* & Stephen A. Johnston†‡

Departments of Medicine and † Biochemistry, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75235-8573, USA

* Present address: Department of Internal Medicine, Barnes Hospital, St Louis, Missouri 63110, USA

‡ To whom correspondence should be addressed

THE acidic transcriptional activation motif functions in all eukaryotes¹⁻⁴, which suggests that it makes contact with some universal component of the transcriptional apparatus. Transcriptional activation by the yeast regulatory protein GAL4 requires an acidic region at its carboxyl terminus. Here we implement a selection scheme to determine whether GAL4 can still function when this C-terminal domain has been deleted. It can, when accompanied by a mutation in the *SUG1* gene which is an essential gene in yeast. Analysis of mutant *SUG1* in combination with various alleles of *GAL4* indicates that *SUG1* acts through a transcriptional pathway that depends on *GAL4*, but requires a region of *GAL4* other than the C-terminal acidic activation domain. The predicted amino-acid sequence of *SUG1* closely resembles that of two human proteins, TBP1 and MSS1, which modulate expression mediated by the human immunodeficiency virus *tat* gene.

GAL4 protein stimulates transcriptional activation of the galactose-regulated (*GAL*) genes by 1,000-fold. This activation requires the C-terminal 28 amino acids of *GAL4* (Fig. 1a). When this acidic region is deleted (*gal4D*), only 3% of wild-type activity remains (Table 1a: *GAL4* *SUG1* versus *gal4D* *SUG1*), even

though the *gal4D* protein is present in normal amounts and its DNA-binding characteristics are unaffected (data not shown). As a strain bearing the *gal4D* gene is unable to grow on galactose medium, we isolated suppressor mutations by selecting for a galactose-growth phenotype. Two complementation groups other than *GAL4* were found. One class of mutants was characterized further and found to have at least a 10-fold higher level of *GAL4*-regulated transcription (Table 1; and northern blot, data not shown). Genetic analysis of this mutant strain demonstrated that the mutation suppressing the *gal4D* lesion was in a previously identified gene, *SUG1* (ref. 5). The original *sug1-1* allele⁵, as well as the 17 that we identified, were recessive. The mutant alleles had no marked effect on the growth rate in a wild-type background and only presented a phenotype in a strain containing the *gal4D* allele grown on galactose.

To test the allelic specificity of *sug1-1*, we combined a number of *gal4* mutants (Fig. 1c), each bearing an alteration in the C-terminal activation domain reducing its activity, with *sug1-1*. As indicated in Table 1a, *sug1-1* clearly suppresses the defect in two (*gal4-3*, *gal4-GN870*) of three of these defective *gal4s*. One of the suppressed alleles (*gal4-GN870*) contains a different truncation from *gal4D*, as well as mutations that eliminate all acidic amino acids in this domain, whereas *gal4-3* has two missense mutations. There is little if any effect of *sug1-1* on the wild-type or *gal4-Cys867* alleles, possibly because they are already strong activators. Thus it seems that the *sug1-1* suppression has the largest effect on weak activators. These results suggest that the *sug1-1* protein is not interacting directly with the C terminus of *GAL4* but partially relieves the requirement for this acidic domain, either by acting up- or downstream in the transcription pathway and/or through another activation function of *GAL4*. This effect is similar to that shown by the mutant form of *GAL11*, *GAL11P* (ref. 6). *GAL11* is a non-essential general transcription factor required for full expression from several transcription factors, including *GAL4* (ref. 7), *PPR1* (ref. 6), and *RAP1* (ref. 8). *GAL11P* can enhance weakly activating derivatives of *GAL4*. *GAL11P*, like *sug1-1*, has little effect

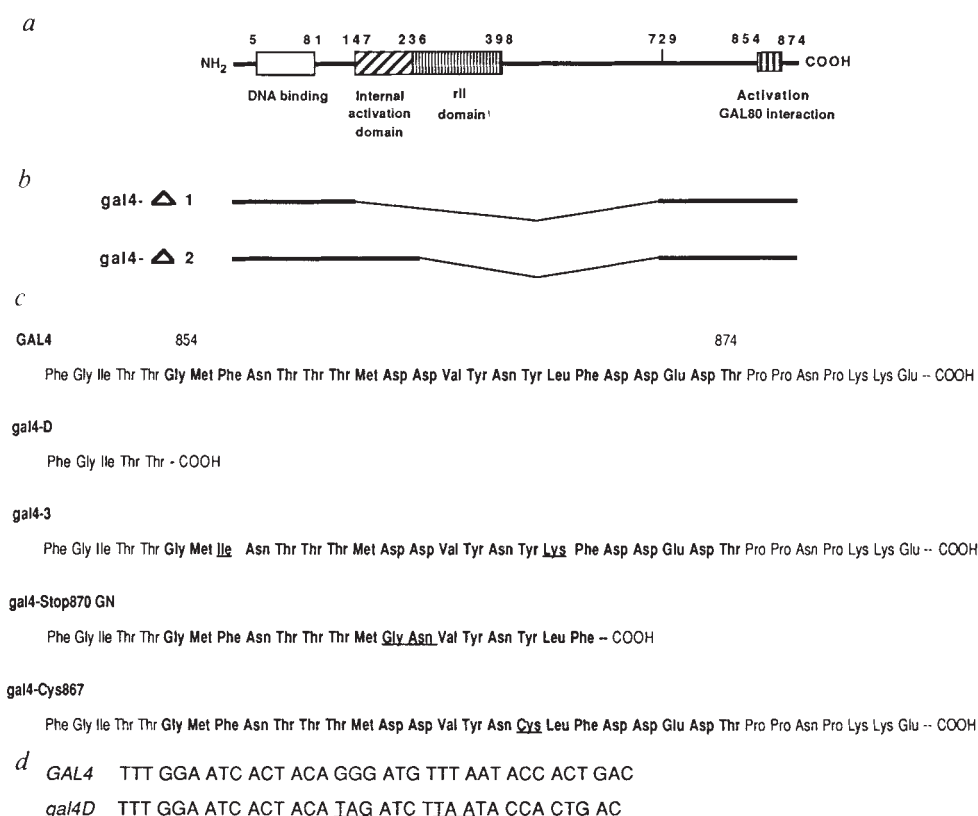


FIG. 1 *GAL4* mutants used in this study will be described in detail elsewhere (K. K. Leuther, J. M. Salmeron and S.A.J., manuscript in preparation): **a**, Entire wild-type *GAL4* protein showing domain organization; **b**, schematic showing positions of the two internal deletion mutants; **c**, amino-acid sequences of the C-terminal region of *GAL4* wild type and mutants. *gal4D* was engineered by insertion of two stop codons and a frameshift. Amino acids in bold represent the activation region, with mutated positions indicated by underlining. **d**, The nucleotide sequence of wild-type *GAL4* and *gal4D* (starting at amino acid 849) (underlined nucleotides indicate stop codons and frameshift).

TABLE 1 Suppression of different *gal4* mutants by *sug1-1*

	Plasmid	β -Galactosidase activity	
		SUG1	sug1
(a)	pSB32	0	0
	pSB32 GAL4	2280	2820
	pSB32 gal4D	72	1500
	pSB32 gal4-3	960	1800
	pSB32 gal4-GN870	240	1380
	pSB32 gal4-Cys867	1800	1860
(b)	pSB32 gal4- Δ 1	30	48
	pSB32 gal4- Δ 2	240	390
	BM 813 gal4 Ser 322-Phe	252	210
	BM 814 gal4 Leu 331-Pro	0	0
	BM 797 gal4 Ser 352-Phe	0	0
	BM 812 gal4 Ser 511-Pro	0	0

β -Galactosidase activities of isogenic strains of yeast YJ0Z, (*gal4^Δgal80^ΔSUG1 GAL1-10lacZ*) and YJ0ZS (*gal4^Δgal80^Δsug1-1 GAL1-10lacZ*) transformed with different activation-deficient mutants of GAL4 carried as *Bam*HI/*Hind*III fragments on the single-copy plasmids pSB32 or pTC3 (BM mutants⁹). Cells were collected from minimal glycerol/lactate cultures at midlog (3×10^7 cells per ml). Extracts were prepared using glass beads, and soluble protein determined using the Pierce BCA method. Assays were done in 1.2 ml 100 mM sodium phosphate, pH 7, 10 mM KCl, 1 mM MgSO₄, 50 mM β -mercaptoethanol containing 20 μ g soluble protein at 30 °C. Values represent nmol *o*-nitrophenyl- β -D-galactopyranoside (ONPG) hydrolysed per min per mg soluble protein. Values are averages of two individual transformants from separate transformations, except for pSB32 GAL4 and pSB32 gal4-D, which are averaged from nine individual transformants. Standard deviations are less than 25%.

on wild-type GAL4. But whereas GAL11 serves as an activator when fused to a DNA-binding domain⁶, SUG1 does not (data not shown).

To determine whether *sug1-1* suppression is specific for C-terminal activation domain mutants, or is effective on all poorly activating forms of GAL4 (by stabilizing the GAL4 protein for example), we tested two internal deletions of GAL4 (gal4 Δ 1 and gal4 Δ 2) (Fig. 1b) and four missense alleles (BM797, 812, 813 and 814; ref. 9) which severely reduce GAL4 activation. None of these internal mutants was rescued (Table 1b). It is interesting to note that whereas gal4 Δ 2 contains the internal activation domain (amino acids 147–235) described in ref. 10 and gal4 Δ 1 does not, neither mutant contains a domain rII (amino acids 232–398) defined by similarity to the GAL4 homologue LAC9 (ref. 11; Fig. 1a). Three of the missense alleles

```

10      20      30      40      50
MTAVTSSNI VLETHESGK PYFEQKIQET ELKIRSKTEN GRRLEAQRNA

60      70      80      90      100
LNDKVRFTKD ELRLIQEPGS YGVEVIKIVS DRKVLVQVQF EGYKIVDVAK

110     120     130     140     150
DINMKDKAS QKVCIRSDSY MLHKVLENKA DPLVSLMVE KVPDSTYDMV

160     170     180     190     200
GGLTKQIKET KEVIELPVKH PELFESLGIA QPKGVILYGF RGTGKILLAR

210     220     230     240     250
AVAHHTDCKF IRVSGAELIQ KYIGEGSRMV RELFVAREH APSIIFMDEI

260     270     280     290     300
DSIGSTRVEG SGGGSEVQR TMLLELNQLD GFETSKNIKI IMATNRLDIL

310     320     330     340     350
DPALLRPGRI DRKIEFPFPS VAAREILIRI HSRKMLNTRG INLRKVAEKM

360     370     380     390     400
NGCSGADVKG VCTEAGMYAL RRRRIHVITQ DFELAVGVM NQNQETAISV

```

FIG. 2 The predicted amino-acid sequence of SUG1 (one-letter code). METHODS. The nucleotide sequence was determined by DNA sequencing of a 1.7 kb *Eco*RI/*Kpn*I fragment inserted into pUC118. The EMBL Data Library accession number is X66400. The ATPase site is underlined.

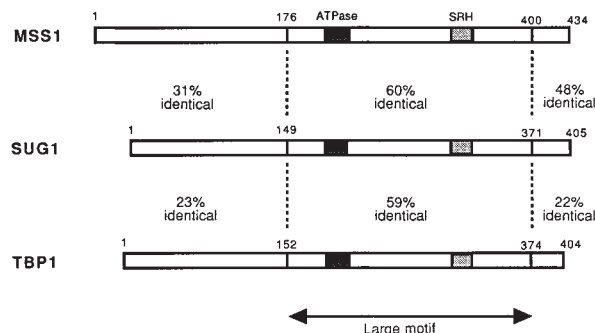


FIG. 3 Comparison of SUG1, MSS1 and TBP1 amino-acid sequences using the Wisconsin Genetics Group program BESTFIT. Indicated is the large motif recently identified in a number of proteins¹³. This motif consists of a consensus ATPase site positioned 88 amino acids from another highly conserved second region of homology (SRH) of unknown function. The two regions are contained within a larger (~220 amino-acid) domain of homology that has been recognized in proteins involved in secretion (SEC18 and NSF)^{15,16}, cell cycle (CDC48)¹⁷, and peroxisomal protein localization (PAS1)¹³. This motif is also found in VCP¹⁸ and its homologue Xlp97 (ref. 19), proteins of unknown function.

(BM797, 813, 814) also lie within rII, but the fourth mutation (BM812) lies outside this region. All four missense mutations abolish the phosphorylation of GAL4 upon induction¹². We conclude that *sug1-1* does not simply suppress all weak GAL4s but requires an internal region of GAL4 protein for its function.

The wild-type *SUG1* gene was cloned and sequenced, and a deletion of one allele created in a diploid strain. When this diploid was sporulated, the spores segregated in a ratio of 2:2, viable to inviable. The disrupted allele was not evident in any of the viable spores, so *SUG1* is presumably essential in yeast. Immunolocalization of a β -galactosidase-SUG1 fusion protein revealed SUG1 to be located in the nucleus (data not shown).

The sequence of the *SUG1* gene reveals a coding region of 405 amino acids (*M_r* 45.2K) (Fig. 2). Examination for structure/function motifs revealed a recently identified motif¹³ found in a number of proteins with different functions. The most striking similarities are between SUG1, TBP1 (ref. 14) and MSS1 (K. Matsumoto, personal communication) (Fig. 3). Both TBP1 and MSS1 were isolated from human libraries and can affect human immunodeficiency virus (HIV) *tat* transactivation in certain cell lines.

The gene we have isolated, SUG1, is essential in yeast and mutations in it can largely relieve the requirement for the C-terminal acidic activation domain in GAL4. The interplay between the *sug1-1* allele and mutations in *GAL4* suggest that *sug1-1* suppression requires an internal region of GAL4. The fact that deletion of the C-terminal activation domain largely inactivates GAL4 indicates that normally SUG1 activation is dependent on this acidic region. The *sug1-1* allele apparently by-passes this dependence. The simplest model is that wild-type SUG1 is a general transcription factor that interacts with GAL4 through both its C terminus and an internal region. The *sug1-1* protein only requires the internal region of GAL4 to act as a co-activator. Regardless of the point of action, it is clear that SUG1 and presumably TBP1 and MSS1 are members of a new family of transcription factors. The isolation of homologues from both yeast and humans may provide insights into cellular factors regulating *tat* activation. □

Received 3 February; accepted 13 May 1992.

1. Fischer, J. A., Giniger, E., Maniatis, T. & Ptashne, M. *Nature* **332**, 853–856 (1988).
2. Ma, J., Przbilla, E., Hu, J., Bogorad, L. & Ptashne, M. *Nature* **334**, 631–633 (1988).
3. Kakidani, H. & Ptashne, M. *Cell* **52**, 161–167 (1988).
4. Webster, N., Jin, J. R., Green, S., Hollis, M. & Chambon, P. *Cell* **52**, 169–178 (1988).
5. Matsumoto, K., Adachi, Y., Toh-E, A. & Oshima, Y. *J. Bact.* **141**, 508–527 (1980).

6. Himmelfarb, H. J., Pearberg, J., Last, D. H. & Ptashne, M. *Cell* **63**, 1299–1309 (1990).
7. Suzuki, Y., Nogi, Y., Abe, A. & Fukasawa, T. *Molec. cell. Biol.* **8**, 4991–4999 (1988).
8. Nishizawa, M., Suzuki, Y., Nogi, Y., Matsumoto, K. & Fukasawa, T. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5373–5377 (1990).
9. Johnston, M. & Dover, J. *Genetics* **120**, 63–74 (1988).
10. Gill, G. & Ptashne, M. *Cell* **51**, 121–126 (1987).
11. Salmeron, J. M. & Johnston, S. A. *Nucleic Acid Res.* **14**, 7767–7781 (1986).
12. Mylin, L. M., Johnston, M. & Hopper, J. E. *Molec. cell. Biol.* **10**, 4623–4629 (1990).
13. Erdmann, R. et al. *Cell* **64**, 499–510 (1991).
14. Nelbock, P., Dillon, P. J., Perkins, A. & Rosen, C. *Science* **248**, 1650–1653 (1990).
15. Eakle, K. A., Bernstein, M. & Emr, S. D. *Molec. cell. Biol.* **8**, 4098–4109 (1988).
16. Block, M. R., Glick, B. S., Wilcox, D. E., Wieland, F. T. & Rothman, J. E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7582–7586 (1988).
17. Fröhlich, K.-U. et al. *J. Cell Biol.* **114**, 443–453 (1991).
18. Koller, K. J. & Brownstein, M. J. *Nature* **325**, 542–545 (1987).
19. Peters, J.-M., Walsh, M. J. & Franke, W. W. *EMBO J.* **9**, 1757–1767 (1990).

ACKNOWLEDGEMENTS. We thank K. Leuther and M. Johnston for *GAL4* mutants; J. True for sequencing the *sgv1-1* allele and M. Goebel for comparison of unpublished protein sequences. This work was supported by grants from the NIH and the Human Frontiers Science Program (S.A.J.), the Perot Family Foundation and the Moss Heart Trust.

New human gene encoding a positive modulator of HIV Tat-mediated transactivation

Hiroshi Shibuya*, Kenji Irie†, Jun Ninomiya-Tsuji†, Mark Goebel‡, Tadatsugu Taniguchi* & Kunihiko Matsumoto‡§

* Institute for Molecular and Cellular Biology, Osaka University, Suita-shi, Osaka 565, Japan

† Department of Molecular Biology, Faculty of Science, Nagoya University, Chikusa-ku, Nagoya 464-01, Japan

‡ Department of Biochemistry and Molecular Biology, The Walther Oncology Center, Indiana University School of Medicine, Indianapolis, Indiana 46202-5122, USA

THE human immunodeficiency virus-1 (HIV-1) protein Tat is a potent activator of virus gene expression^{1,2}. Tat functions through a sequence known as TAR, located immediately downstream of the transcription start site in the long terminal repeat^{3–5}. Several observations suggest that cellular factors cooperate with Tat in the overall transactivating process. We have isolated a human complementary DNA from the new gene *MSS1*, which may encode such a cellular factor, by transcomplementation of a yeast *sgv1*[–] mutant. The *MSS1* protein shares 42% sequence identity with the human TBP-1 protein, which binds Tat *in vitro* and suppresses Tat-mediated transactivation *in vivo* (ref. 6). We report here that the levels of HIV activation by Tat correlate with endogenous levels of *MSS1* messenger RNA. Furthermore, we provide evidence that expression of *MSS1* enhances the Tat-mediated transactivation. Our results suggest that *MSS1* has a key role in activation of HIV genes regulated by Tat.

We previously identified the *SGV1* gene of the budding yeast *Saccharomyces cerevisiae* which encodes a CDC28/CDC2-related kinase and seems to regulate G1 cyclins⁷. An *sgv1* mutation confers cold-sensitive growth. To isolate mammalian genes that regulate the yeast *SGV1*-mediated pathway, we screened an expression library for human cDNAs that can bypass a defective *SGV1* function. The human cDNA library was produced from HeLa cells, and was cloned into the yeast expression vector pNV11 in which the *TDH3* (glyceraldehyde-3-phosphate dehydrogenase) promoter and terminator flank the cDNA cloning site⁸. This library was transformed into a *S. cerevisiae* cold-sensitive mutant strain containing *sgv1* and *ura3* mutations. A total of 5×10^5 Ura⁺ colonies were obtained and screened for their ability to grow at the restrictive temperature of 17 °C. One clone was isolated that showed a plasmid-linked complementing activity. This gene was named *MSS1* (mammalian suppressor of *sgv1*).

The *MSS1* cDNA was found to be 1.5 kilobases (kb) long, which is consistent with the observed mRNA size (about 1.6 kb; see Fig. 3). Its nucleotide sequence and deduced amino-acid sequence are shown in Fig. 1. The encoded protein is 433 amino acids long, with an estimated relative molecular mass of 48,633. The predicted *MSS1* amino-acid sequence contains one putative nucleotide-binding domain and reveals strong similarity to human TBP-1, a 404-amino-acid protein that binds to HIV-1 Tat and negatively regulates Tat function⁶; *MSS1* and TBP-1 share 42% identity over 374 amino acids (Fig. 2).

The expression pattern of the *MSS1* gene was analysed by northern blotting. Total cellular RNA was isolated from various cell lines and hybridized with radiolabelled *MSS1* cDNA. A unique 1.6-kb transcript was detected in human Jurkat (T cell), Chinese hamster ovary (CHO), and murine P19 (embryonic carcinoma) cells, reflecting the conservation of *MSS1* in different species. Similar results were also obtained in HeLa and NIH3T3 cells (data not shown). RNA expression was highest in Jurkat and lowest in P19 cells (Fig. 3).

The structural homology between the *MSS1* and TBP-1 proteins suggested a possible involvement of *MSS1* in the regulation of HIV-1 Tat function. To examine this possibility, the *MSS1* cDNA was placed under control of the human elongation factor promoter in a mammalian expression plasmid (pEF-*MSS1*). When pEF-*MSS1* was cotransfected with a Tat-expressing plasmid (pSRα-Tat) and an HIV long terminal repeat (LTR) indicator plasmid (pBencat; in which the HIV LTR controls

```

1  CCATTGTGCTCTAAAGGGAAGGCTGCTGTGTAATCATTAAAGGCGGAGGCTTTGGAGCTGCTAAATGCCGATTACCT
1  M P D Y L
81  CGGTGCCGCTACCGCGAAGACCAAGAGGATGAGAAGGACGACAGCCATCCGAGCTCTGGATGAGGGGGATTCGCT
6  G A D Q R K T K E D K D D K P I R A L D E G D I A L
161  TGTGTAACCTTATGTCGAGCAGCTTACTTCAGGAGATCAAGCAGTGAAGATGACATTCAGCACTTCGTAAGAAA
33  L K T Y G Q S T Y S R Q I K Q V R D D I Q Q L L K K
241  ATTAATGAGCTCACTGGTATTAAAGAATCTGACACTGGCTGGCCGCCAGCAGCACTGGGATTTGGTGCAGATAAGCA
59  I N E L T G I K E S D T G L A P P A L T G W D L A A D K A Q
321  GACACTCCAGGTGAACAGCTTTACAGCTTGCACAGGTGACAAAGATATCAATGCTGATTCGGGAGCCCAAAATACA
86  T L Q G E Q P L Q V A R C T T K I N A D S E D P F K Y I
401  TTATCAACGTAAGCAGTTTGCCAAAGTTTGGTGGAGCTTATGATGACCTGAGTCACTGAGTCACTGAAGAGGGATG
113  I N V K Q F A K F V D L S D Q V A P T D I E E G M
481  AGATGGGCGTGATAGAAATAATATCAATTCACATTCATTCGCTTCTAAGATGAGCCAGCAGTATACCATGATGCA
139  R V G V D R N K Y Q I H I P L P P K I C P T V T M M Q
561  GGTGGAAGACAACTGATGACATACAGTGTGTTGGTGGCTGTGAAGGAGCAAGATTGAGAACTGAGAGAGTGTG
166  V E E K P D V T Y S D V G G C K E Q I E K L R E V V E
641  AAACCCCATCTTCTCCAGAGAGGTTTGTGAACCTTGGCATTGAGCTCCCAAGGGCGCTGCTCTTTGGTCCACC
193  T P L L H P E R F V P P K G V L L F G P
721  GGTACAGGCAAGCAGCTCTGTGGCGGGGAGTGTCTATCGGAGTGTGCTGCTTCAATCGAGTTATGGATCGAGCT
219  G T G K T L C A R A V A N R T D A C F I R V I G S E L
801  TGTCAGAAATACCTCGGTGAGGGGGCTGGAATGTTGCTGAGCTTTTGAAGTGGCAGAGCAAAAAGCGTGGCTTA
246  V Q T Y Y V G E G A R M V R E L F E M A R T K K A C L I
881  TCTTCTTGATGAATGATGCTATTGAGGGGGCTGTTTGTGATGATGCTGCTGAGGTGACAAATGAATGACAGAAACA
273  F F D E I D A I C G A R T D D G A G G N E V Q R T
961  ATGTGTGAGATGATCACTGAGTGTGTTGATGCTCGAGGAGATATTAAGTGTGAGTGGCAGTCAAGAGCTGTA
299  M L E L I N Q L D G F D P R G N I K V L M A T N R P C
1041  TACTTTGGATCCAGCAGTGTGAGGCGGAGAGATGGATAGAAAAATGAATTTAGCTTGCCGATCAGAGGCTCGGA
326  T L D P A L M R P G R L D R K I E F S L P D L E G R T
1121  CCCACATATTGAAGTATCAGCTTCAATGAGGTGTTGAAGAGATATCAGATTTGAATGTTAGCAGCAGTGTGCA
353  H I F K I H A R S M S V E R F E L L A R L C P
1201  AATGAGCACTGCTGAGATTAGAAGCGTCTGACAGAGGCTGTATGTTTTCATTCAGAGACAGCAAGAAATTTCTAT
179  N S T G A E I R S V C T E A G M F A I R A R R F I A D
1281  CGAGAAGGATTTCTTGAAGCTGTAAAGGCTTATTAAGTCTTATGCAAAATTAATTAATTAATTAATTAATTAAT
406  E K D F L E A V N K V I K R Y A K F I A T F R Y N T Y
1361  ACAACTGACCGCTGAGGCTTTCAAGTGAAAAATTTAAATGGGAATCTAAATTTATATAGATTTTAAATACCAATTC
433  N *
1441  ATAAACAAATAAATGGCTTCAACTTTAGACCAATGG

```

FIG. 1 The nucleotide sequence of the *MSS1* cDNA and the predicted amino-acid sequence of the encoded protein (single-letter code). The deduced amino-acid sequence of the putative *MSS1* protein is shown below the nucleotide sequence starting with the first methionine codon in the open reading frame. The sequence data reported will appear on the EMBL Nucleotide Sequence Database under the accession number D11094.

METHODS. The *S. cerevisiae* strain *sgv1 ura3* was transformed with a HeLa cDNA library. Transformants were screened for improved growth at 17 °C. One clone was isolated that showed a plasmid-linked phenotype. Plasmid DNA from this clone was transferred into *Escherichia coli*. Restriction endonuclease fragments of the cDNA insert were cloned into phage vector M13mp18 and M13mp19 and sequenced by the dideoxy chain termination method.

§ To whom correspondence should be addressed.

MSS1	1	MPDYLGAQDRKTEKEKDDKPIRALDEGDIALKTYGQSTYSRQIKQEDDIDQLLRKIN
TBP1	1	MSTETIQRTLRDSEIKIMKSEVLRLTHEIEMKDKK
MSS1	61	ELTGKESDTGAPPAWDLAADKQTQSEQPLVARCTKIINADSEDPKYIINVKQFA
TBP1	40	ENSEKIKVNTKL--PML--VSNVIELLDVDPNDDEEDGANILDSQKRGKCAVLKSTSQ
MSS1	120	KFVVDLSQAPTDIEEHRVGVORNNKQTHIPLPKIPLPTIMKVEEDVDVTSVGG
TBP1	96	TYFLPVLIGLDAEKLKPLDGLVGNKDSQILETLPTFYSRKAMEDEPTEQKSDTGG
MSS1	180	CKEIKTEKREVVETLLRLRRLNGLGEPFRGVILGPPGPGTKTQARAVNRRDALTIR
TBP1	156	LDKIQJQLIVEIVLNNHKKRNLGLIQPPKGVIMYSPPGTGKTJLARAFLACQKTELK
MSS1	240	VISELVCKYVDSARMVRELFEVPRTRKACLTFFDEIDATGGAFTDGGAGSDNEVQRTM
TBP1	216	LAPQQLVCMFTDCKRLDQDAALKEAPSTLEHDEIDATGTRKEDSEKAGSDREVQRTM
MSS1	300	LELLNOLDGEFIRGNIKVLMETNAPDLDPAIRPGRDRKTEFSIDLDLRTKFKTHA
TBP1	276	LELLNOLDGEFIRGNIKVLMETNAPDLDPAIRPGRDRKTEFSIDLDLRTKFKTHA
MSS1	360	RSSVDEIRIRFELLARLCPNSTCAEIRSVCHQAGMIRARRKIAPEKFLAVNKKIKS
TBP1	336	RSSVDEIRIRFELLARLCPNSTCAEIRSVCHQAGMIRARRKIAPEKFLAVNKKIKS
MSS1	420	YAKFSATPRYMTYN
TBP1	396	KKANLQYYA

FIG. 2 Alignment of the MSS1 amino-acid sequence with that of human TBP-1. Identical residues are boxed. Consensus sequences for ATP binding are marked with asterisks.

the expression of a chloramphenicol acetyltransferase (CAT) reporter gene⁹ into P19 cells, Tat-mediated expression of the CAT gene was markedly stimulated (Fig. 4a). In CHO cells, a modest increase in the expression of the CAT gene was consistently observed. By contrast, no significant increase in the expression of the HIV-LTR-CAT was observed in Jurkat cells. The levels of HIV activation by Tat correlate with the endogenous levels of MSS1 mRNA. This observation may in part explain the different effects of MSS1 expression on Tat function in different cell lines. In fact, it is possible that with respect to the Tat-mediated transactivation, the MSS1 protein is already expressed in saturating amounts in Jurkat cells, but not in P19. To investigate whether the levels of MSS1 expression would affect the function of another viral activator, we cotransfected HTLV-1 (human T-cell lymphotropic virus type 1) LTR-CAT indicator plasmid (pLTRcat; in which the HTLV-1 LTR controls the expression of CAT)¹⁰ and a tax-1 (the transactivator protein of HTLV-1; refs 11, 12) expression plasmid (pCDS) (ref. 13) into P19 cells. As shown in Fig. 4b, efficient activation of HTLV-1 LTR by tax-1 was observed in P19 cell lines under the same condition where Tat activated HIV-LTR very weakly. Hence, these results suggest that the tax-1 function is not significantly affected by MSS1 expression.

Here we have provided evidence that expression of MSS1 stimulates Tat-mediated activation. Because MSS1 is present in all cell lines tested, it may also be required. In fact, MSS1 seems to function as a positive modulator of Tat. This is in contrast to TBP-1, which seems to function as a negative regulator of

FIG. 4 Effect of MSS1 expression on Tat- or tax-1-mediated activation. Plasmid pEF-MSS1, expressing MSS1 under the control of the human elongation factor (EF) promoter, was cotransfected with the indicated expression plasmid DNAs into Jurkat, CHO or P19 cells. a, The reporter plasmid, pBencat which contains HIV-LTR-CAT (ref. 9), was cotransfected with a Tat expression vector, pSRα-Tat. b, The reporter plasmid, pLTRcat which contains HTLV-1 LTR-CAT (ref. 10), was cotransfected with a tax-1 expression vector, pCDS (ref. 13). The transfection experiments were repeated at least three times and the results were reproducible. In the case of CHO cells, the percent conversions of CAT were 38.0 ± 0.7 and 33.4 ± 0.6 as s.e.m. of four separate experiments for the cotransfection of pSRα-Tat with or without pEF-MSS1, respectively.

METHODS. The MSS1 expression vector, pEF-MSS1, contains the MSS1 coding sequence abutted downstream to the EF promoter/enhancer sequences¹⁷ linked to the vector CDM8 (ref. 18). pSRα-Tat was generated by cleaving

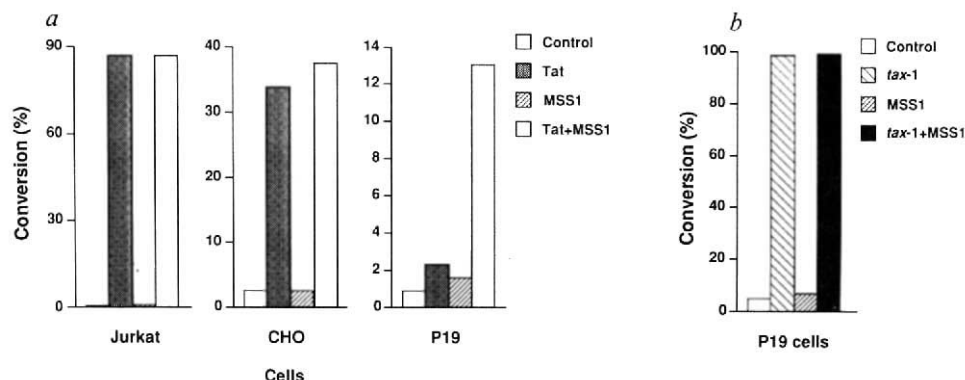


FIG. 3 Northern blotting analysis of MSS1 gene expression. The MSS1 cDNA probe was hybridized to 10 µg total cellular RNA prepared from the cell lines indicated. An overnight exposure of the autoradiogram is shown.

METHODS. Total cellular RNA was prepared by the guanidine thiocyanate method¹⁵. RNAs were northern blotted from 1% formaldehyde gels. Probes were labelled with [α -³²P]dCTP using a Multiprime labelling kit (Amersham) and hybridized as previously described¹⁶.

Tat (ref. 6). In view of the notable structural similarity between MSS1 and TBP-1, an intriguing possibility exists, namely that the two proteins compete with each other to regulate the interaction with the transcriptional complex with the HIV gene, thereby influencing gene expression in both directions. Although the mechanism for MSS1 function has yet to be determined, one can speculate that the Tat and MSS1 proteins form part of a transcriptional initiation or elongation complex. The ability of MSS1 to promote Tat-mediated transactivation indicates that it has a key role in transactivation and contributes to the establishment of latent HIV infection. Therefore, an understanding of the exact mechanism and role for MSS1 action should provide insight toward a means of controlling this process.

Genetic screening of mammalian cDNA in yeast is a powerful tool for the isolation of novel components of mammalian cells¹⁴. But in the case of MSS1, it is still unclear how human MSS1 can bypass the yeast SGV1 function. A yeast homologue of MSS1, *SUG1*, has also been identified; the *SUG1* gene product shows 49% identity with human MSS1 and acts through a transcription pathway that requires an internal region of the GAL4 activator (S. Johnston, personal communication). This suggests that *SUG1* may act in a functionally similar manner

to human MSS1 in yeast. It will be interesting to test whether SUG1 affects SGV1 function. □

Received 17 February; accepted 30 April 1992.

1. Arya, S. K., Gio, C., Josephs, S. J. & Wong-Staal, F. *Science* **229**, 69–73 (1985).
2. Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Science* **228**, 1430–1434 (1985).
3. Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. *Cell* **41**, 813–823 (1985).
4. Hauber, J. & Cullen, B. R. *J. Virol.* **62**, 673–679 (1988).
5. Garcia, J. A. *et al.* *EMBO J.* **8**, 765–778 (1989).
6. Nelbock, P., Dillon, P. J., Perkins, A. & Rosen, C. A. *Science* **248**, 1650–1653 (1990).
7. Irie, K., Nomoto, S., Miyajima, I. & Matsumoto, K. *Cell* **65**, 785–795 (1991).
8. Ninomiya-Tsuji, J., Nomoto, S., Yasuda, H., Reed, S. I. & Matsumoto, K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 9006–9010 (1991).
9. Gendelman, H. E. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 9759–9763 (1986).
10. Kobayashi, N. & Hatanaka, M. *Cancer Res.* **1**, 64–95 (1986).
11. Sodroski, J. G., Rosen, C. A. & Haseltine, W. A. *Science* **225**, 381–385 (1984).
12. Fujisawa, J., Seiki, M., Kiyokawa, T. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2277–2281 (1985).
13. Shibuya, H., Yoneyama, M. & Taniguchi, T. *Int. Immun.* **1**, 43–49 (1989).
14. Lee, M. & Nurse, P. *Nature* **327**, 31–35 (1987).
15. Shibuya, H. *et al.* *Nucleic Acids Res.* **18**, 3697–3703 (1990).
16. Harada, H. *et al.* *Cell* **63**, 303–312 (1990).
17. Kim, D. W., Uetsuki, T., Kajiro, Y., Yamaguchi, N. & Sugano, S. *Gene* **91**, 217–223 (1990).
18. Seed, B. *Nature* **329**, 840–842 (1987).
19. Takebe, Y. *et al.* *Molec. cell. Biol.* **8**, 466–472 (1988).
20. Adachi, A. *et al.* *J. Virol.* **59**, 284–291 (1986).

ACKNOWLEDGEMENTS. We thank S. Nomoto, R. Ruggieri and T. Mizukami for discussion and S. Johnston for communicating information before publication. This work was supported by the Ministry of Education, Science and Culture of Japan (K.M. and T.T.), the Mitsubishi Foundation (K.M.) and an NIH grant (M.G.).

Processing pathways for presentation of cytosolic antigen to MHC class II-restricted T cells

Mauro S. Malnati*, Merce Marti†, Timothy LaVaute*, Dolores Jaraquemada†, William Biddison‡, Robert DeMars§ & Eric O. Long*||

* Laboratory of Immunogenetics, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 12441 Parklawn Drive, Rockville, Maryland 20852, USA

† Immunology Unit, Hospital Germans Trias i Pujol, PO Box 72, 08916 Badalona, Spain

‡ Neuroimmunology Branch, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland 20892, USA

§ Laboratory of Genetics, University of Wisconsin, Madison, Wisconsin 53706, USA

ANTIGENS presented to CD4⁺ T cells derive primarily from exogenous proteins that are processed into peptides capable of binding to class II major histocompatibility complex (MHC) molecules in an endocytic compartment^{1–4}. In contrast, antigens presented to CD8⁺ T cells derive mostly from proteins processed in the cytosol, and peptide loading onto class I MHC molecules in an early exocytic compartment is dependent on a transporter for antigen presentation encoded in the class II MHC region^{5–11}. Endogenous cytosolic antigen can also be presented by class II molecules^{12,13}. Here we show that, unlike class I-restricted recognition of antigen, HLA-DR1-restricted recognition of cytosolic antigen occurs in mutant cells without a transporter for antigen presentation. In contrast, DR1-restricted recognition of a short cytosolic peptide is dependent on such a transporter. Thus helper T-cell epitopes can be generated from cytosolic antigens by several mechanisms, one of which is distinct from the classical class I pathway.

To test whether class II-restricted presentation of endogenously processed matrix protein of influenza A virus^{12,13} was representative of other cytosolic proteins, a cytosolic form of the H3 haemagglutinin transmembrane antigen of influenza

A virus was inserted into a recombinant vaccinia virus (Vac-cytoH3). A protein with the expected relative molecular mass of 62,000 (62K) and a half-life of about 2 h was detected with a pool of anti-H3 monoclonal antibodies in B cells infected with Vac-cytoH3 (Fig. 1). The relative stability of these cytoH3 molecules contrasts with the instability of a similar cytosolic form of the H1 haemagglutinin reported previously¹⁴. By comparison, cells infected with Vac-H3 expressed the fully glycosylated form of H3 with M_r 83K (Fig. 1).

CytoH3 expressed in cells infected with Vac-cytoH3 was presented to DR1-restricted T cells at least as efficiently as H3 expressed in Vac-H3 infected cells (Fig. 2). Presentation of both H3 and cytoH3 was inhibited by 80 μ M chloroquine (Fig. 2a), as had been observed with presentation of endogenous matrix protein¹³. Vaccinia virus preparations inactivated by ultraviolet light did not sensitize target cells (Fig. 2b), demonstrating a

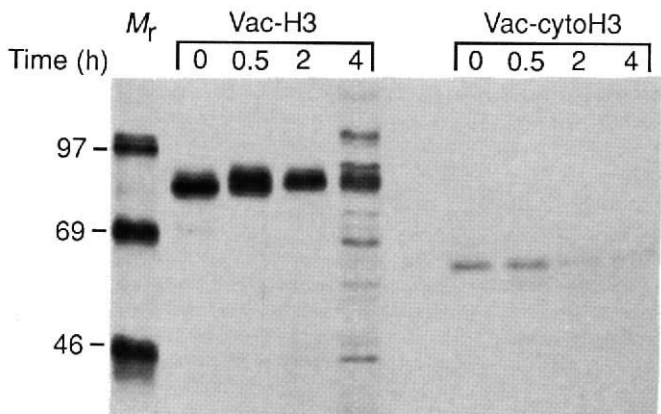


FIG. 1 Synthesis of H3 haemagglutinin and of a signal sequence-deleted H3 haemagglutinin in cells infected with recombinant vaccinia viruses. The Epstein-Barr virus-transformed B-cell line .45 (ref. 22) was infected with recombinant vaccinia viruses encoding the H3 haemagglutinin (Vac-H3) or a cytosolic form of H3 haemagglutinin (Vac-cytoH3), labelled with ³⁵S-methionine for 20 min and chased for the indicated time in hours. Cells were lysed and H3 molecules were immunoprecipitated with a pool of monoclonal antibodies specific for H3 (a gift from R. Webster), separated on an 8% polyacrylamide-SDS gel and autoradiographed. $M_r \times 10^{-3}$ for ¹⁴C-labelled markers (Amersham) indicated on the left.

METHODS. Vac-H3 (ref. 23) was a gift from J. Bennink. Vac-cytoH3 was constructed as follows. Plasmid p3X31A⁺ (a gift from M. J. Gething) encoding the H3 haemagglutinin of influenza A virus was digested with *SalI* (filled with dCTP and dTTP with Klenow polymerase) and *EaeI* to produce a cDNA fragment starting just after the coding region for the H3 signal sequence. Plasmid pEL1 was constructed from plasmid pSC11-ss (ref. 24) by inserting a synthetic oligonucleotide that provided *SmaI* and *BglII* cloning sites and retained the *SalI* site. Plasmid pEL1 digested with *SalI* (blunt-ended with Klenow polymerase) and *BglII* (filled with dATP and dGTP) was ligated with the *EaeI-SalI* cDNA fragment using synthetic oligonucleotide adaptors that provided 29 bp of the H3 5' untranslated region and the first two codons of H3. The resulting construct encoding the first two amino acids of the H3 precursor linked to amino acid 17 and the rest of H3 (Met-Lys-Gly-Gln-) was verified by sequencing. The *SalI* site of the cDNA fragment and the *BglII* site of pEL1 were made compatible by partial filling as already described. pEL1-cytoH3 was used to generate a recombinant vaccinia virus as described²⁵. For infections, 2×10^7 .45 cells were incubated with 60 PFU per cell of sucrose gradient-purified vaccinia viruses at 4 °C for 1 h in 2 ml serum-free medium, and further incubated on a rotator for 1 h at 37 °C. Fetal calf serum was added to a final concentration of 5% and the cells were incubated for another hour on a rotator at 37 °C. The infected cells were washed, starved in methionine-free medium for 1 h, and labelled with 0.5 mCi ³⁵S-methionine (ICN) for 20 min and chased with 1 mM methionine. Cells were lysed in 2% Triton X-100, 10 mM Tris-HCl, 150 mM NaCl in the presence of 0.5 mM PMSF and 5 mM iodoacetamide at 4 °C for 1 h. The supernatant was first pre-cleared with control mouse ascites (Sigma) and a polyclonal anti-mouse IgG antiserum (Calbiochem) bound to protein A-Sepharose (Sigma), before adding anti-H3 monoclonal antibodies. The specific immunoprecipitation was carried out using standard procedures.

|| To whom correspondence should be addressed.

dependence on newly synthesized H3 or cytoH3 proteins for presentation. Furthermore, the bulk of newly synthesized cytoH3 was not transported through the exocytic pathway because it remained unglycosylated (Fig. 1). But the possibility that T cells detect processing of rare cytoH3 molecules translocated into the exocytic pathway cannot be ruled out. Cells infected with a vaccinia virus expressing the influenza virus matrix protein (Vac-M1) were not recognized by the H3-specific T cells (Fig. 2c).

To test whether processing of endogenous cytoH3 for presentation to DR1-restricted T cells may follow the class I pathway, a recombinant vaccinia virus was constructed that expresses a short cytosolic peptide containing the 12-amino-acid epitope of H3 flanked by two amino acids at either end (Vac-miniH3). Such 'mini-gene' constructs have been used to generate peptides for class I-restricted presentation¹⁵⁻¹⁷. Although presentation of endogenous miniH3 to the DR1-restricted T cell line C1.6 was either very low (Fig. 2c) or undetectable in other experiments (not shown), another DR1-restricted T cell specific for H3 (E1.9) was able to recognize cells infected with Vac-miniH3 (Fig. 3a). As was observed with both H3 and cytoH3, presentation of endogenous miniH3 to E1.9 was blocked by 80 μ M chloroquine (Fig. 3a). Under the same conditions, chloroquine did not affect class I-restricted presentation of endogenous matrix¹³.

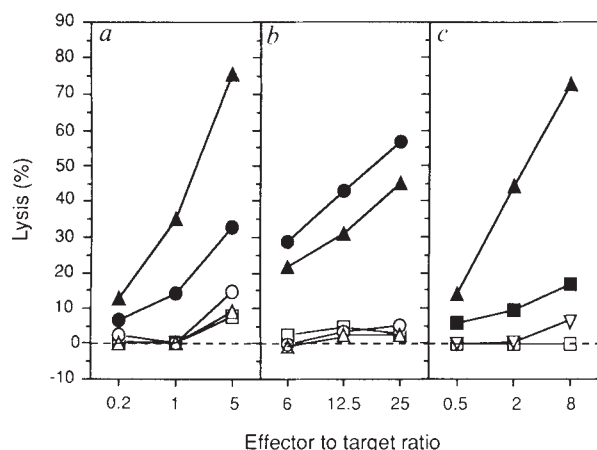


FIG. 2 Presentation of newly synthesized H3 haemagglutinin and cytosolic H3 haemagglutinin to DR1-restricted T cells. Lysis of the B cell line .45 was measured with the DR1-restricted H3-specific T cell line C1.6 (ref. 26) at the indicated effector to target ratios. Cells were either uninfected (open squares), infected with 20 PFU per cell Vac-H3 (circles), or 20 PFU per cell Vac-cytoH3 (triangles). Cells were infected with vaccinia viruses either in the absence (filled symbols) or the presence of chloroquine (open symbols). Chloroquine at 80 μ M was present during the entire infection, and was reduced to 10 μ M during the CTL assay. *b*, Cells were infected with either intact vaccinia viruses (filled symbols), or viruses inactivated with 1,200 μ W cm^{-2} of ultraviolet light for 15 min (open symbols). *c*, Cells were infected with 10 PFU per cell Vac-miniH3 (filled squares), or 30 PFU per cell Vac-M1 (ref. 27) (inverted triangles).

METHODS. 5×10^5 .45 cells were infected in 0.5 ml serum-free medium with sucrose gradient-purified vaccinia viruses as described in Fig. 1 legend, except that the infection was allowed to proceed at 37 $^{\circ}\text{C}$ for a total of 4 h, and that after 1 h at 37 $^{\circ}\text{C}$, cells were diluted with 1 ml of medium to a final fetal calf serum concentration of 5%. Cells were labelled with 50 μCi sodium [^{51}Cr]chromate for 1 h in 0.2 ml, washed, counted and plated at 5×10^3 cells in V-bottom 96-well plates containing previously aliquoted effector cells. Vac-miniH3 was constructed as follows. A synthetic oligonucleotide encoding the sequence MAPKYVKQNTLKLAIL was ligated into *Sal*I-*Bgl*II-digested plasmid pEL1. The 5' end sequence of the synthetic oligonucleotide was TCGACCATGG, thus providing a good translation initiation signal²⁸. The sequence of the oligonucleotide inserted into pEL1 was verified by sequencing in both orientations. pEL1-miniH3 was used to generate a recombinant vaccinia virus as described²⁵. Experiments shown in *a*, *b* and *c* were not done at the same time. The different effector to target ratios used in these separate experiments indicate the varying level of activity of the T-cell line C1.6 at these different times.

The difference between E1.9 and C1.6 cells in their ability to recognize miniH3 may be due to a higher affinity for antigen of E1.9, or to distinct fine specificities. Half-maximal lysis by E1.9 was obtained at 1/4 the concentration of synthetic peptide required for half-maximal lysis by C1.6 (not shown). Peptide truncations showed, however, that the minimal epitope required by E1.9 was longer than that required by C1.6 (not shown).

Mixing experiments demonstrated that presentation of cytoH3 and of miniH3 did not involve molecules released from the infected cells and processed through the exogenous pathway. Cells infected with Vac-cytoH3 (Fig. 3b) or Vac-miniH3 (Fig. 3c) did not sensitize co-cultured cells for lysis by E1.9 T cells, implying that processing was endogenous.

To avoid the difficulty of interpreting results obtained with inhibitors such as chloroquine, a genetic approach was chosen to test whether the class I pathway was involved in the class II-restricted presentation of cytoH3 and miniH3. Mutant .134 (ref. 18), derived from .45 cells, has a defect in subunit I of the transporter for antigen presentation (TAP1), which is necessary for proper class I assembly and class I-restricted presentation of endogenous antigen^{11,19}. In agreement with a recent report¹¹, mutant .134 was defective in presentation of Vac-M1 to HLA-A2-restricted T cells (data not shown). Furthermore, processing of exogenous influenza virus matrix and haemagglutinin antigens by mutant .134 for DR1-restricted presentation was comparable to processing by the parental .45 cells²⁰. It was thus possible to test whether TAP1 was necessary for class II-restricted presentation of endogenous antigen.

Mutant .134 infected with Vac-cytoH3 was efficiently lysed by E1.9 T cells (Fig. 4a, b). This suggested either that delivery of peptides to class II molecules in an early exocytic

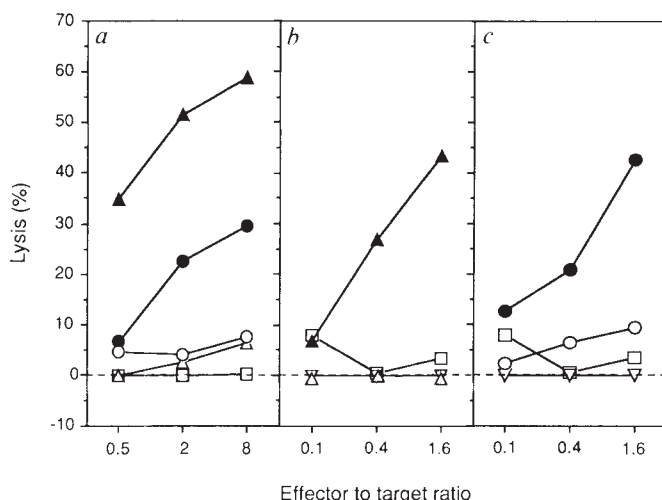


FIG. 3 Endogenous processing of cytosolic H3 and of a cytosolic H3 peptide for presentation to DR1-restricted T cells. Lysis of .45 cells was assayed with the DR1-restricted H3-specific T cell line E1.9 (ref. 26) at the indicated effector to target ratios. Uninfected cells (squares) or cells infected with Vac-M1 (inverted triangles) were included as negative controls. *a*, Cells were infected with 10 PFU per cell Vac-cytoH3 (triangles) or Vac-miniH3 (circles) in the absence (filled symbols) or the presence (open symbols) of chloroquine as described for Fig. 2. *b*, .45 cells were labelled with sodium [^{51}Cr]chromate and subsequently infected with Vac-M1 as described for Fig. 2. Infected cells were washed at the end of the incubation in serum-free medium and mixed with an equal number of unlabelled cells infected with Vac-cytoH3 (also washed at the end of the incubation in serum-free medium). The mixed cells (triangles) were incubated for another 4 h at 37 $^{\circ}$ on a rotator before the CTL assay carried out as described in Fig. 2 legend. Conversely, labelled cells infected with Vac-cytoH3 were mixed with unlabelled cells infected with Vac-M1 (filled triangles). *c*, Labelled cells infected with Vac-M1 and mixed with unlabelled cells infected with Vac-miniH3 (circles), or labelled cells infected with Vac-miniH3 and mixed with unlabelled cells infected with Vac-M1 (filled circles) were assayed as in *b*.

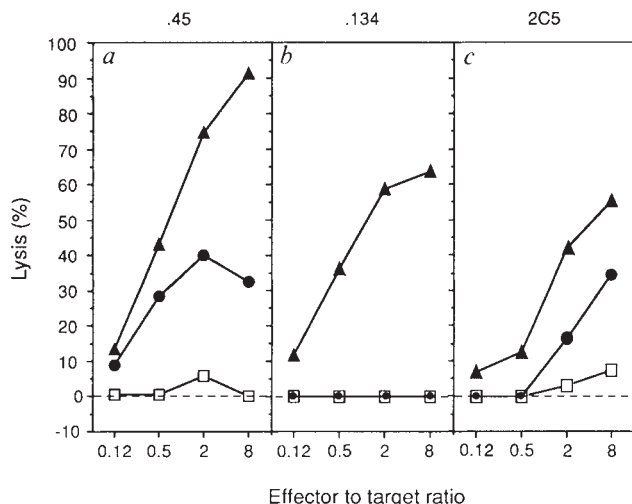


FIG. 4 A functional TAP is required for DR1-restricted presentation of a cytosolic H3 peptide but not for presentation of cytosolic H3. Lysis of uninfected (squares) or infected (filled symbols) cells was tested with the DR1-restricted H3-specific T-cell line E1.9 at the indicated effector to target ratios. Cells were infected with Vac-cytoH3 (triangles) or Vac-miniH3 (circles) as described for Fig. 2. *a*, Epstein-Barr virus-B cell line .45. *b*, TAP1-deficient mutant .134 derived from .45. *c*, Transfectant 2C5 of mutant .134 with restored TAP1 expression.

compartment occurs in a TAP-independent manner, or that a distinct pathway exists for the delivery of cytosolic antigen to a class II compartment. It is unlikely that cytoH3 presentation is mediated by the TAP2 subunit alone as mutant .134 cannot present the class I-restricted M1 epitope. Furthermore, mutant .134 infected with Vac-miniH3 was not recognized by E1.9 cells (Fig. 4b). The defect in presentation of endogenous miniH3 was clearly due to the lack of functional TAP and not to a secondary mutation, because transfectant .134 cells expressing a complementary DNA for TAP1 (clone 2C5)¹¹ were lysed by E1.9 after infection with Vac-cytoH3 and Vac-miniH3 (Fig. 4c), even though the overall efficiency of lysis of transfectant 2C5 was somewhat reduced compared with .45 cells. Presentation of an endogenous cytosolic peptide by class II molecules is thus dependent on a functional TAP1. Therefore, presentation of endogenous cytoH3 in the absence of TAP1 probably does not involve translocation of short peptides into an early exocytic compartment.

These results define distinct pathways for the presentation of endogenous antigen by class II molecules. The first is similar to the class I pathway in that it uses short cytosolic peptides in a TAP-dependent process. But it is distinct from the class I pathway in its sensitivity to chloroquine. In addition, whereas presentation of cytosolic peptides by class I molecules is extraordinarily efficient, it is not yet clear how efficient or widespread this pathway may be for class II-restricted presentation. The second pathway is distinct from the class I pathway in that it does not require TAP1 function. Furthermore, it may not involve short peptides, but rather the delivery of larger cytosolic molecules to an endosomal/lysosomal compartment for processing, as was suggested earlier in a study on endogenous processing of cytosolic matrix protein^{13,21}.

It is now clear that antigenic epitopes recognized by CD4⁺ T cells can be generated from processing of cytosolic proteins by several mechanisms. A potential use of such mechanisms is to accelerate the generation of T-cell help by class II-positive cells that are virus-infected. Although some viruses require endocytosis for infection, many others infect cells by fusion at the cell surface, delivering viral proteins directly into the cytosol. In the latter case, such as with human immunodeficiency virus, endogenous processing of cytosolic proteins for class II-restricted presentation may be an important defence mechanism. An important question raised by our findings is whether class II-restricted presentation of cytosolic proteins extends to self proteins in normal cells and to what extent such presentation may influence T-cell repertoire and tolerance. □

Received 28 February; accepted 11 May 1992.

1. Unanue, E. R. & Allen, P. M. *Science* **236**, 551-557 (1987).
2. Morrison, L. A., Lukacher, A. E., Braciale, V. L., Fan, D. P. & Braciale, T. J. *J. exp. Med.* **163**, 903-921 (1986).
3. Rudensky, A. Y., Preston-Hurlburt, P., Hong, S.-C., Barlow, A. & Janeway, C. A. Jr *Nature* **353**, 622-627 (1991).
4. Germain, R. N. & Hendrix, L. R. *Nature* **353**, 134-139 (1991).
5. Townsend, A. & Bodmer, H. A. *Rev. Immun.* **7**, 601-624 (1989).
6. Cox, J. H., Bennis, J. R. & Yewdell, J. W. *J. exp. Med.* **174**, 1629-1637 (1991).
7. Monaco, J. J., Cho, S. & Attaya, M. *Science* **250**, 1723-1726 (1990).
8. Deverson, E. V. *et al. Nature* **348**, 738-741 (1990).
9. Trowsdale, J. *et al. Nature* **348**, 741-744 (1990).
10. Spies, T. *et al. Nature* **348**, 744-747 (1990).
11. Spies, T. *et al. Nature* **355**, 644-646 (1992).
12. Nuchtern, J. G., Biddison, W. E. & Klausner, R. D. *Nature* **343**, 74-76 (1990).
13. Jaraquemada, D., Marti, M. & Long, E. O. *J. exp. Med.* **172**, 947-954 (1990).
14. Townsend, A. R. M., Bastin, J., Gould, K. & Brownlee, G. G. *Nature* **324**, 575-577 (1986).
15. Whittton, J. L. & Oldstone, M. B. A. *J. exp. Med.* **170**, 1033-1038 (1989).
16. Gould, K., Cossins, J., Bastin, J., Brownlee, G. G. & Townsend, A. *J. exp. Med.* **170**, 1051-1056 (1989).
17. Sweetser, M. T., Morrison, L. A., Braciale, V. L. & Braciale, T. J. *Nature* **342**, 180-182 (1989).
18. DeMars, R., Chang, C. C., Shaw, S., Reitnauer, P. J. & Sondel, P. M. *Hum. Immun.* **11**, 77-97 (1984).
19. Spies, T. & DeMars, R. *Nature* **351**, 323-324 (1991).
20. Ceman-Bogner, S., Rudersdorf, R., Long, E. O. & DeMars, R. *J. Immun.* (in the press).
21. Long, E. O. *New Bio.* **4**, 274-282 (1992).
22. Kavathas, P., Bach, F. H. & DeMars, R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4251-4255 (1980).
23. Gould, K. G., Scotney, H., Townsend, A. R. M., Bastin, J. & Brownlee, G. G. *J. exp. Med.* **166**, 693-701 (1987).
24. Earl, P. L., Koenig, S. & Moss, B. *J. Virol.* **65**, 31-41 (1991).
25. Chakrabarti, S., Brechling, K. & Moss, B. *Molec. cell. Biol.* **5**, 3403-3409 (1985).
26. Karp, D. R. & Long, E. O. *J. exp. Med.* **175**, 415-424 (1992).
27. Smith, G. L., Levin, J. Z., Palese, P. & Moss, B. *Virology* **160**, 336-345 (1987).
28. Kozak, M. *Nucleic Acids Res.* **15**, 8125-8128 (1987).

ACKNOWLEDGEMENTS. We thank M. Weston for technical assistance, P. Roche for comments on the manuscript, the Biological Resources Branch of NIAID for peptide synthesis, J. Bennis, P. Earl, M. J. Gething, B. Moss and R. Webster for antibodies, plasmids and viruses.

ERRATUM

Myosin head movements are synchronous with the elementary force-generating process in muscle

Malcolm Irving, Vincenzo Lombardi,
Gabriella Piazzesi & Michael A. Ferenczi

Nature **357**, 156-158 (1992).

THE seventh sentence in the description of Fig. 1b on page 156 in this Letter should begin "The intensity before the step was $92 \pm 4\%$ of that in a tetanus without imposed steps", not $9 \pm 4\%$ as printed.

Working in the chemical industry

Colleen Shannon

Many companies are looking beyond the academic records of applicants these days, and are searching for potential managers. A few of those taken on can look forward to truly international careers.

A FEW years ago, profit figures for the western chemical industry were reaching record levels. The nightmare of the early 1980s was over. Nearly a decade of plant closures, job cuts and the general blood-letting prescribed by the recession had left the industry leaner and meaner, if not a little dizzy.

Confident plans were taking shape, too. Many companies wanted to move away from traditional bulk chemicals, where low profit margins and cyclical demand are seen as the norm, into higher-value products. The industry spent money on new factories, new research, and new recruits.

The honeymoon was short. The chemical industry, which sells most of its products to the manufacturing sector, has been hit hard by the recession of the 1990s. In the past few years, nearly every major chemical company has announced some sort of cost-cutting initiative; nearly everywhere, the workforce has grown smaller.

ICI, Britain's largest chemical company (and one of the world's top five) reduced its workforce by some 9 per cent in 1991, cutting almost 12,000 jobs worldwide. In 1989, member companies surveyed by the UK Chemical Industries Association hired some 1,160 graduates. (About 80 per cent of these recruits held first degrees, while the rest held advanced degrees.) By 1991, these employers were planning to take on 650 graduate recruits — a reduction of about 44 per cent.

Even in the Netherlands — where management is traditionally homegrown and new recruits are definitely seen as long-term prospects — recruitment is down at major companies like DSM. The German chemical giant, Bayer, is planning 3,000 job cuts worldwide.

Across the Atlantic, Dow Chemical has added several major businesses to its portfolio, but total employee numbers have remained flat. For the first time in 25 years, Du Pont's US recruiters are bypassing the campus circuit.

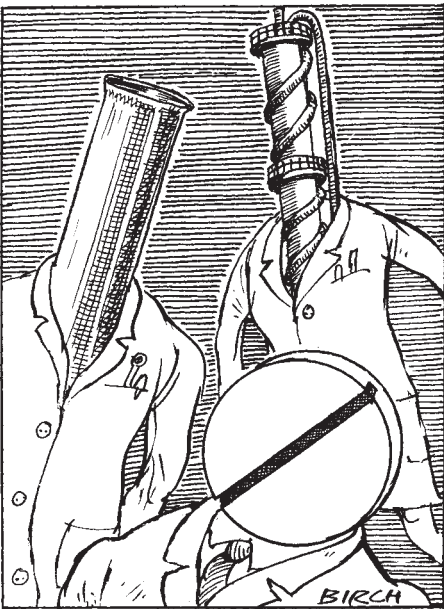
The situation is not hopeless, however. In the United States, the American Chemical Society (ACS) keeps track of employment conditions with an annual membership survey. In March of last year, 1.6 per cent of respondents were unemployed. This compares with a rate of nearly 7 per cent among the entire US workforce at that time. The figure is also lower than chemist unemployment rates

recorded in 1983, the peak year of the last recession. However, this year's data are likely to show an upturn in chemist unemployment, according to ACS staff.

There are some sunny spots, such as the pharmaceuticals sector, where profits have remained strong. Even where the industry is cutting back, many managers still perceive a shortage of truly competent chemists and engineers. And many companies, such as ICI, have decided that it is prudent to keep on recruiting through the hard times.

Industry option

Who wants to work for the chemical industry, anyway? For one thing, there is



the environmental problem. Do graduates balk at the idea of working for an industry that is often seen, in spite of its attempts to reach the public, as the big green bogeyman?

In Carl Bauer's experience, the answer is no. Bauer, who has been in the business for some 30 years, is now director of personnel resources for Dow Chemical in the United States. He has "hired hundreds of kids". Yet during that time, only one objection has crossed his desk. Besides, he says, people working in the industry are uniquely placed to offer real solutions to environmental problems. Curiously enough, he believes this very argument is attracting more women into the traditionally male stronghold of chemical engineering.

For some reason, women seem to be

migrating towards the pharmaceutical industry, too, according to David Rogers, recruitment manager at SmithKline Beecham. About half of the company's chemical engineering recruits are now female, he says. This is in a field where women represent only about 13 per cent of UK graduates.

The fact is that most chemists will eventually end up in industry, or at least somewhere on the business side of chemistry. Nearly 60 per cent of full-time UK chemists work in industry or commerce, according to the latest membership remuneration survey of the Royal Society of Chemistry (RSC). Figures from the ACS show a similar trend.

The old truism that salaries are higher in industry is borne out by the ACS survey figures. At the PhD level, the median salary for US academic chemists is just under \$48,000 a year. Their peers in industry can expect \$63,000 a year.

Starting salaries are respectable, too. At Dow Chemical, a PhD chemist should collect about \$50,000 in the first year of employment. At Du Pont, a chemist or biologist with a bachelor's degree starts off at around \$32,000 a year.

In the UK, the difference is cut finer. Until chemists reach the age of 30 or so, median academic salaries hover at around £15,000 a year, while the industry figure is £18,600 a year. However, by the time these two groups of chemists reach retirement age, the gap between them narrows.

Industrial starting salaries are not as generous in the UK, either. At ICI, typical graduates can look for £14,000 a year, while new PhD chemists will start at about £15,500. The figures look brighter in the research-oriented pharmaceutical industry. At SmithKline Beecham, for example, PhD chemists start off at £18,000 to £20,000 a year.

Chemists in Germany seem to fare better. At Bayer, typical PhD graduates start out at Dm90,000 a year — about twice the going rate at ICI. And in the Netherlands, a chemist with similar qualifications would start out at DSM at Fls85,000, again, a significantly higher figure.

Even if the money is right, is the chemical industry the place for someone who really loves research chemistry? At ICI, only one in seven graduates works in research and development. At Dow, the number is one in five, but that still

leaves a lot of people working in production, marketing, sales and other support roles.

Recruiters admit that, even in the laboratory, industry chemists are held to a commercially focused agenda. This is especially true in the initial stages of a researcher's career. But many employers do try to give scientists time to follow up some of their own ideas. This is the practice at SmithKline Beecham, where management does not want researchers to "feel too confined".

After a long and distinguished career, the most talented industry scientists can attain a certain state of academic freedom. For instance, becoming one of Du Pont's elite research fellows means that "you can virtually do whatever you want to do" in terms of research, says the company's senior college relations supervisor, Richard Koffenberger.

Many major chemical companies now have parallel career paths, which allow scientists to stay in research and development without sacrificing advancements in status and pay. Take, for instance, ICI's scientific ladder programme, which has been in place since 1965. The system resembles a pyramid, with hundreds of locally appointed research associates at the base, and just eight international leaders at the top. These senior research associates earn just as much money as their counterparts in management, an ICI spokeswoman explains. The company is currently considering a relaunch, to increase employee awareness of the programme.

There is no doubt that industry can often provide first-rate research facilities. And these days, many academics find that they actually have little time for research. Science must often be squeezed in between grant proposals, the search for graduate students to staff the laboratory on a shoestring, and a round of administrative duties. "For some of these people, this is not a good life", one recruiter says. People considering a career in an academic institution might find out "it's not the life they envisioned", he warns.

Going global

For the very ambitious and the very capable, the chemical industry can provide the key to a truly international career. Going abroad almost always means working for the company in your own country for a while, and working your way up. Some companies are very reluctant to move staff around the globe, while others would like to do it more often. But in general, the industry has a strong tradition of internationalism, and every major company has operations abroad.

Paul Schindler exemplifies that tradition. A Frenchman, he came to Britain

in 1965 to work for ICI as a mathematician. He spoke little English, and says his job interview was basically conducted on a blackboard. Schindler later moved up through the ranks in sales and marketing, and has since held senior posts in Belgium, France and China; he is now based in Singapore, and heads the company's Asia-Pacific business. He believes that ICI is stronger for its tradition of tolerance, which runs across cultural boundaries. By putting a French manager in charge of a major Asian business, this British company "is showing a certain strength of mind", Schindler observes.

He also notes that the Asia-Pacific region now accounts for a major share of the industry's worldwide growth. Most of ICI's recruiting in the region is local for now, but the company wants to place more expatriates in the longer term, Schindler says.

Those expatriates will have to be able to cope with the frenetic pace, the "horrendous" travelling (Singapore to Hong Kong is more than three hours by plane) and the immense cultural differences, Schindler says. Language skills are a prerequisite. Finally, he believes that family needs have to be considered very seriously.

Those who do make their way to Asia — and decide they like it — will find themselves living in a Renaissance of sorts, Schindler says. "There is a dynamic, a drive, that you don't find anywhere else in the world".

Getting there

In spite of the competition engendered by hard economic times, managers in the industry are not uniformly impressed by the quality of today's chemistry graduates. The word on the ground is that the best graduates are still hard to find, according to Martin Hunt, secretary for professional affairs at the RSC.

"The best" do not necessarily come from certain universities, either. What this industry needs is people who can make things work. At SmithKline Beecham, David Rogers sums up his company's primary requirement: "We need practical people for practical jobs."

In addition to a strong academic background, his company is looking for some evidence of hands-on ability. It helps if students have done a practical thesis at university, he says. The company is not impressed by paper projects.

At Dow, Carl Bauer agrees. Students who can find a summer job in the industry should do so, he urges. In fact, the company has set up its own special training scheme. Students on Dow's intern "co-op" plan alternate between campus and the company during their university studies. This means that a US undergraduate degree will take five

years to complete, rather than the usual four, but Bauer thinks the experience makes the extra time worthwhile.

While there is always room for biologists, environmental scientists and statisticians, most technical vacancies in the industry are filled by chemists or engineers. At some companies, computer scientists are in demand; at others, information technology jobs are disappearing fast in the cost savings drive.

Certain specialists are in high demand. At the US Chemical Manufacturers Association, a spokesman observes that a "cascade of regulations" has been "washing over this industry in recent years". The industry needs environment-minded engineers who can make sure that facilities are in compliance with these new laws, he observes.

The needs, of course, vary with the company. But in general, many are on the lookout for chemical engineers. Du Pont, for one, obviously puts a premium on these skills; starting pay for engineering graduates outstrips pay for other science graduates by about a fifth. Many companies need more analytical chemists, and candidates with a strong background in organic synthesis are likely to feel welcome, especially in the pharmaceutical industry.

Advice from Martin Hunt at the RSC: "Learn to sell yourself on paper." Hunt, who runs the RSC's modest employment referral service, sees hundreds of application forms, and not all of them do justice to their authors. "Some people think that because you are in science, you don't have to worry about these things", he says.

Even the most meticulous applicant, though, is likely to have some trouble after the age of 45, in Hunt's experience. "As they get older, it gets a lot tougher."

Perhaps that has something to do with his final point. Companies these days are looking beyond the academic record, Hunt believes. When applicants come in for interview, the manager is likely to be wondering where, five or ten years down the line, they will fit in. At ICI, for example, recruitment marketing manager Bob Pugh says the company is definitely looking for potential managers.

Sometimes, the most important qualities can be difficult to pin down precisely. Jos van Ruyven, head of corporate recruitment for DSM, believes that new employees have to find their place in the company pretty quickly. "If you're unable to do that, I think you'll have trouble". And at Bayer, applicants are not only expected to have excellent academic credentials — they should also have the ability to "think in networks".

Colleen Shannon is a trade journalist specializing in the chemical industry.